THE FUNCTION OF ALYREF ON UAP56-MEDIATED R-LOOP RESOLUTION

by

Kathryn Primrose Banks, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biochemistry
May 2020

Committee Members:

Xiaoyu Xue, Chair

Karen Lewis

L. Kevin Lewis

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

I would like to thank Dr. Xue for his willingness to teach me the skills needed to successfully perform thorough research, and the opportunity to work in his laboratory. I would also like to thank my committee members, Dr. Karen Lewis, and Dr. Kevin Lewis for their support and guidance through this process. I would also like to thank the members of the Xue lab during my time at Texas State for contributing to such a healthy, collaborative learning environment. Most of all, I would like to thank my family and loved ones for the continuous support; I truly could not have accomplished this without all of you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

R-loops arise during transcription, and serve important roles including class switch recombination, protection of genes from DNA methylation at GC rich promoters, and promotion of transcription termination, but they can also cause replication stress and double-strand breaks if they are allowed to accumulate and are not resolved by the cell's innate processes of R-loop resolution, which greatly contributes to genomic instability. Cells do have methods of resolving persistent R-loops, but the mechanism of how these DNA replication and repair factors prevent R-loop accumulation remains unclear. Mutations in the TREX complex cause accumulation of co-transcriptional R-loops, and components of TREX, UAP56, and ALYREF, are of particular interest. UAP56 depletion in humans has been linked to a strong genomic instability phenotype, demonstrated in part by preliminary unpublished data showing that depletion of UAP56 results in an increased level of R-loops in *HeLa* cells. ALYREF associates with UAP56, and has been found overexpressed in cancerous tissue. These implications prompt the exploration of the function of ALYREF on UAP56 mediated R-loop resolution activity. Our data provided proof that UAP56 has R-loop dissociation activity, emphasized that ALYREF had a direct effect on the DNA/RNA helicase and R-loop dissociation activities of UAP56, and explored the basis of interaction between the two proteins through ALYREF variants.

I. INTRODUCTION

Genomic Instability and R-Loops

Genomic instability can have devastating effects. One of these effects is cancer. Now generally referred by the CDC as malignant neoplasms, cancer has been one of the top three leading causes of death in the United States since 1927 [1]. Trailing just behind heart disease, cancer continues to be one of the leading causes of death in the US, with an estimated 630,000 deaths projected for 2020 [2]. In addition to cancer, many other diseases arise from genomic instability, such as Fanconi anemia, ALS, and Huntington's Disease [3]. Thus, gaining more insight into how genomic instability occurs could have a great impact on the health of current and future generations. Through this project, we sought more information on the genomic instability that is derived from R-loops, which arise in vivo during transcription.

Transcription is the process of synthesizing RNA using DNA sequences as templates. R-loops arise during transcription; they are unusual three-strand nucleic acid structures formed by an RNA:DNA hybrid plus a displaced single strand DNA (ssDNA). R-loops serve important roles in various biological processes, including class switch recombination, protection of genes from DNA methylation at GC rich promoters, and promotion of transcription termination [4-7]. However, aberrant R-loop accumulation, if not resolved properly and promptly, can be pathological, and is linked to neurodegenerative diseases (ALS, Fragile X syndrome, and Friedreich's ataxia) and cancer [3, 8, 9]. While functional R-loops generally resolve with the cell's innate processes of R-loop resolution, lingering R-loops can cause many issues, such as replication stress, and double-strand breaks (DSBs), which are thought to be one of the

most harmful DNA lesions to occur because of their contribution to genomic instability [10] (Figure 1).

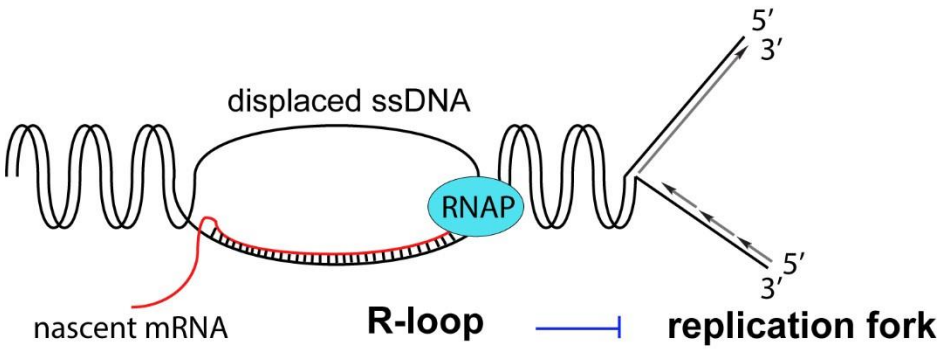

Figure 1. The co-transcriptional R-loop and subsequent fork stalling. The RNA polymerase moves towards the 3' end, forming an mRNA strand that is annealed to its parent strand while displacing the complementary DNA strand. If the nascent mRNA threads back to the template strand, it will prevent transcription from taking place again by blocking the progress of the RNAP. This stress can cause replication fork collapse, which can lead to a DSB.

Mechanism of R-Loop Formation

One generally accepted mechanism of R-loop formation is called the "thread-back" model. In this model, the nascent RNA strand exits RNAP, and then "threads back" to the template DNA strand to displace one strand of the duplex DNA [5] (Figure 2). Other models do exist, such as the "extended hybrid" model, but it is not as widely accepted or as well described as the thread-back model. In humans, current data indicates that RNA processing factors play a role in preventing R-loop accumulation, such as SETX, an RNA/DNA helicase, and AQR, an ATP-dependent RNA helicase within the spliceosome [11-13]. These are consistent with the thread-back model, which further emphasizes the importance of a properly assembled RNA and RNA binding protein complex in order to prevent R-loop formation [14].

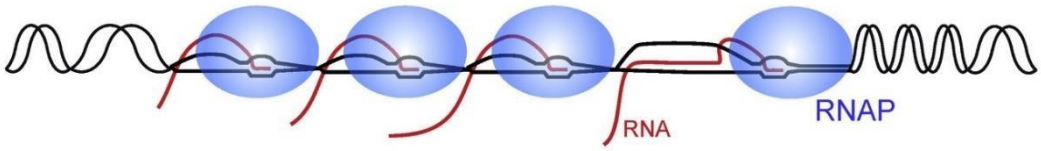

Figure 2. The thread-back model. The mRNA strand has threaded back to the parent DNA strand, creating an R-loop and preventing the RNA polymerase from moving along the dsDNA to proceed with transcription [5].

In vivo, generally, during transcription in areas of high GC content or highly expressed genes are where R-loops occur. However, transcription is necessary for all DNA-containing cells, so it is important to note that R-loops can be functional. They do arise naturally as intermediates through some necessary cellular processes, including mitochondrial DNA replication and immunoglobulin class switching [5]. Yeast cells were the first cells used to provide evidence indicating that R-loops are a source of genome instability [15]. These mutant cells were developed without specific RNA biogenesis and processing factors such as the THO complex, and the data were later further supported when these factors were identified to prevent R-loop accumulation *in vitro* [11, 15]. This accumulation of structures can induce a DSB, and although there is not a defined way these breaks occur, there are a few mechanistic proposals. One possibility is that the stable binding of the RNA:DNA hybrid leaves the single-strand DNA more exposed and susceptible to DNA damage than it would be if it were paired with its complementary strand [16]. Another possibility is that the R-loop causes a sort of "road-block" and prevents the DNA polymerase from proceeding with elongation, thus causing a DSB [5].

R-Loop Resolution Mechanisms

Cells have developed a number of factors to remove the deleterious and persistent R-loops. Most importantly, R-loops can be removed by RNase H enzymes. Through ribonuclease action, the RNA portion of the RNA:DNA hybrid can be targeted and cleaved in a sequence-independent manner. Secondly, R-loops can be resolved by a number of DNA/RNA helicases both *in vitro* and *in vivo* [17-19]. Another R-loop accumulation prevention measure, DNA replication-associated repair, is possible through the action of BRCA1, BRCA2, Fanconi anemia DNA repair and other protecting factors of the replication fork (RF) [20-22]. The mechanism of how these DNA replication and repair factors prevent R-loop accumulation remains unclear. It is noteworthy that persistent R-loops interfere with the progression of DNA replication forks, thus causing fork arrest and collapse, which, again, leads to DSB formation [5, 23].

Importantly, mutations in the TREX (transcription and export) complex, a conserved protein complex involved in transcription and mRNA export, cause accumulation of co-transcriptional R-loops. This links RNA metabolic functions to R-loop processing [24]. In yeast, homologues Yra1 and Sub2 both associate with the heterotetrameric THO complex, which functions in transcription similarly to the human THO complex. The coordination of Yra1 and Sub2 with the THO complex make up what is known as the TREX complex. In yeast, the TREX complex may function as a mediator in coupling transcription to mRNA export, as data has shown it is recruited to transcribed genes where it travels with the DNA polymerase as it carries out elongation [24]. THO subunits are a core component of TREX, which allow it to function as nuclear exporter of mRNAs. Additionally, the TREX complex is conserved in humans.

In humans, components of TREX include the DEAD box type helicase, UAP56, and adaptor mRNA binding protein, ALYREF. In mammals, the TREX complex contains components Thoc1, Thoc2, Thoc7, Thoc5, and Thoc6. Thoc3 is not a stable component of the THO complex for multiple species, although it does interact, and could be considered a THO-interacting factor (Table 1) [25]. The mechanism of how the TREX complex functions specifically in preventing R-loop accumulation remains unknown.

Table 1. Some components of the TREX complex. a: does not consistently act as a stable component of the TREX complex, but is included because of its interaction with the other components.

Yeast	Mammals
THO components	
Hpr1	Thoc1
Tho2	Thoc2
Thp2	
Mft1	Thoc7
	Thoc5
	Thoc6
Tex1	Thoc3$_a$
DEAD-box type helicase	
Sub2	UAP56
Adaptor mRNA binding protein	
Yra1	ALYREF

UAP56 and ALYREF in R-Loop Resolution

UAP56 is a DEAD box type RNA helicase that associates with Thoc2 in the TREX complex. UAP56 depletion in humans has been linked to a strong genomic

instability phenotype [26]. Preliminary unpublished data from our collaborator, Dr. Andres Aguilera, demonstrated that the depletion of UAP56 results in an increased level of R-loops in HeLa cells, indicating UAP56 plays an important role in the resolution of R-loops in vivo. DRIPc-seq with detection by immunofluorescence and DNA:RNA immunoprecipitation provide proof that UAP56 is recruited across the human genome to transcriptional regions to prevent R-loops. Our biochemical data suggests that purified UAP56 possesses RNA/DNA helicase activity in addition to RNA/RNA helicase activity, as well as a robust unwinding activity on a synthetic structure mimicking R-loops. These data are consistent with the hypothesis that low-frequency, but highly problematic lingering co-transcriptional R loops are prevented in part by recruitment of UAP56. In addition to the THO complex, UAP56 also associates with ALYREF, which is the mRNA export adaptor protein in the mammal TREX complex. ALYREF associates with UAP56 through its C-terminal UAP56 binding motif and N-terminus, and stimulates its ATPase and helicase function [27] (Figure 3).

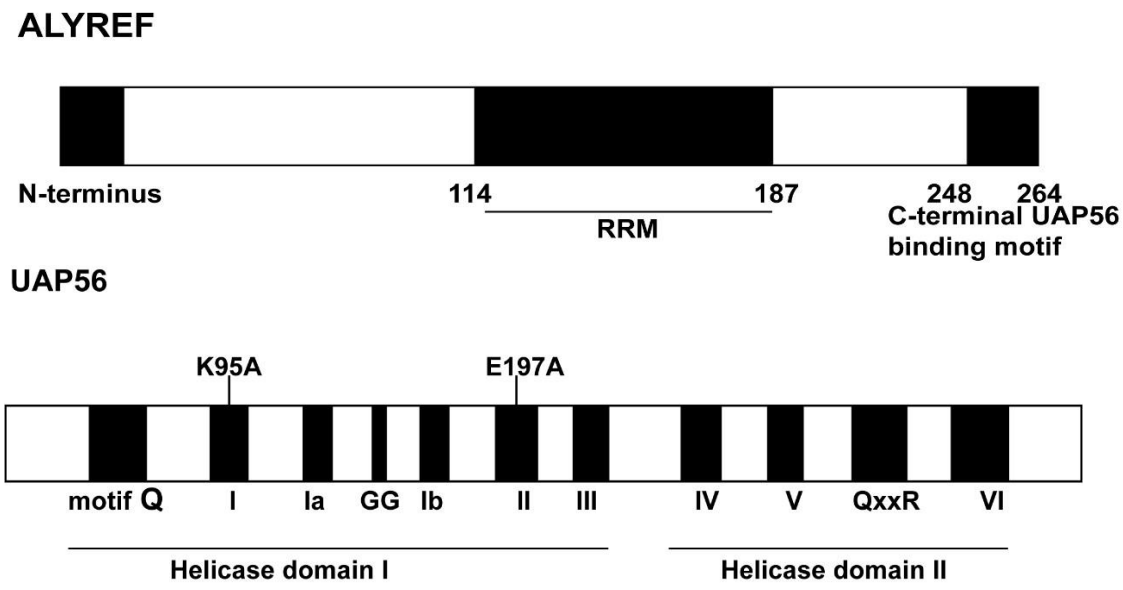

Figure 3. Schematic for relevant domains of UAP56 and ALYREF. Note the C-terminal UAP56 binding motif is 16 amino acids long, with the RNA recognition motif comprising over 25% of the protein. The mutations of UAP56 used in our experiments reside in helicase domain I.

Curiously, ALYREF has been found overexpressed in various cancerous tissues [28, 29]. In cancer cells depleted in ALYREF, there is a significant reduction of the cells' metastatic capacity, and proliferation [29]. Given that UAP56 is recruited across the human genome to prevent R-loops, and that ALYREF is able to associate with UAP56, we were interested to gain insight on the effect of ALYREF on the R-loop resolution capabilities of UAP56. The reduction of cancer cells' metastatic capacity, and proliferation in ALYREF depleted cells also suggests that targeting TREX and its component ALYREF may have therapeutic benefits. Furthermore, the investigation on the regulation of the R-loop resolving capabilities of UAP56 by ALYREF may provide new potential cancer therapy strategies by interrupting ALYREF and UAP56's interactions.

In this study, we purified UAP56 wild type protein as well as its helicase dead mutants K95A and E197A [30], and human ALYREF-WT protein. We first confirmed the strong interaction between UAP56 and ALYREF-WT by pull-down assays, and examined the effect of ALYREF on UAP56's R-loop resolution activity. We characterized the ability of UAP56 to resolve R-loop flaps, and provided proof that these functions of UAP56 are dependent on its ATP hydrolysis using the specific helicase dead mutants E197A and K95A. Data indicates that Lys-95 is crucial to the binding of ATP, while Glu-197 plays a key role in ATP hydrolysis, therefore alanine substitutions were utilized to abolish helicase activity by preventing ATPase activity [31]. To explore the interaction between ALYREF and UAP56, we synthesized two mutants, ALYREF-ΔC16 and ALYREF-ΔNC. Mutants ALYREF-ΔRRM, ALYREF-10E, and ALYREF-17E were also synthesized to test if ALYREF's RNA binding capabilities are important for R-loop resolution mediated by UAP56. These mutants allowed us to further explore the basis of interaction between the two proteins, and assess to what extent the regulation of the R-loop resolving capabilities of UAP56 is mediated by ALYREF.

II. MATERIALS AND METHODS

Reagents

Invitrogen 1 Kb plus DNA Ladder, Invitrogen Blue Juice, Invitrogen 6x-His HRP conjugated antibody, Pierce Protease Inhibitor tablets, and Casein Ultrapure were purchased from Thermo Fisher Scientific (Fair Lawn, NJ). 6X loading dye, T4 Poly Nucleotide Kinase (PNK), and T4 PNK 10X buffer were purchased from New England Biolabs (Beverly, MA). Phenylmethylsulfonyl fluoride (PMSF) was purchased from MP Biomedicals (Santa Ana, CA). Bis-Acrylamide 40% solution, tetramethylethylenediamine (TEMED), Coomassie Brilliant Blue R-250, boric acid, potassium phosphate monobasic, potassium chloride, sodium phosphate dibasic heptahydrate, chloramphenicol, ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), Tween20, yeast extract, potassium hydroxide, sodium hydroxide, tris base, sodium chloride, ethidium bromide, ammonium persulfate, proteinase K, methanol, glacial acetic acid, ethanol, and hydrochloric acid were all purchased from Fisher Scientific (Waltham, MA). Tryptone RPI, glycerol ACS Grade, and kanamycin monosulfate were purchased from Research Products International (Mount Prospect, IL). Acros imidazole, and phosphoric acid were purchased from Thermo Fisher Scientific (Fair Lawn, NJ). Igepal CA-630 was purchased from MP Biomedicals LLC (Irvine, CA). Stained and unstained ladder standards, Mini-PROTEAN TGX gels, Micro Bio-Spin P-6 columns, and molecular biology grade agar were all bought from Bio-Rad Laboratories (Hercules, CA). P^{32}-γ-ATP was bought from PerkinElmer Inc. (Waltham, MA). QIAprep Spin Miniprep Kits were purchased from Qiagen (Hilden, Germany). Nickel Sepharose High Performance resin, SP Sepharose Fastflow resin, and Q Sepharose Fastflow resin

were bought from GE Healthcare (Princeton, NJ).

Equipment

The slab gel dryer model GD 2000 was purchased from Hoefer Inc. (Holliston, MA). Corning Spin-X UF concentrators were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). The Fraction Collector F9-C, AKTA Pure HPLC, HisTrap FF 5 mL column, HisTrap desalting column, Mono S column, 50 mL Superloop, and Superdex 75 Increase 10/300 were all purchased from GE Healthcare (Princeton, NJ). The Trans Blot Turbo transfer system, 50 mL plastic columns, mini size transfer stacks, Mini-Protean tetra system, Protein II xi Cell, ChemiDoc Imaging System, UltraRocker rocking platform, C1000 Touch Thermal Cycler, Screen Eraser-K, and PharosFX Plus Molecular Imager were all purchased from Bio-Rad Laboratories (Hercules, CA). The QSonica Sonicator Model CL-334 was purchased from QSonica LLC (Newtown, CT). The Centrifuge 5810 R 15 amp version, New Brunswick Innova 44 Incubator Shaker, and Centrifuge 5424 were purchased from Eppendorf (Hamburg, Germany). The Isotemp Incubator, Isotemp GPD 10, and Mettler Toledo SevenGo Duo pH meter were purchased from Fisher Scientific (Waltham, MA). The Sorvall LYNX 6000 Superspeed centrifuge, Quantity One Nanodrop, Barnstead Genpure Pro, heating blocks and heaters, and Sorvall Legend micro 17 centrifuge were all purchased from Thermo Fisher Scientific (Fair Lawn, NJ). The Vacuum System was purchased from Vacuubrand, Inc. (Essex, CT).

Origin of Plasmids

The human ALYREF-WT-pET24b was a gift from Stuart A. Wilson at the University of Sheffield. ALYREF-ΔRRM-pET24a, ALYREF-ΔC16-pET24a, ALYREFΔNC-pET24a, ALYREF-10E-pET24a, and ALYREF-17E-pET24a were all

purchased from Gene Universal. UAP56-WT-pGEX-KG, UAP56-E197A-pGEX-KG, and UAP56K95A-pGEX-KG were gifts from Rui Zhao at University of Colorado.

Sequences of Expressed Proteins

Human ALYREF-WT

MPDSAPAMADKMDMSLDDIIKLNRSQRGGRGGGRGRGRAGSQGGRGGGAQAA
ARVNRGGGPIRNRPAIARGAAGGGGRNRPAPYSRPKQLPDKWQHDLFDSGFGG
GAGVETGGKLLVSNLDFGVSDADIQELFAEFGTLKKAAVHYDRSGRSLGTADVH
FERKADALKAMKQYNGVPLDGRPMNIQLVTSQIDAQRRPAQSVNRGGMTRNRG
AGGFGGGGGTRRGTRGGARGRGRGAGRNSKQQLSAEELDAQLDAYNARMDTS

Human ALYREF-ΔC16

This mutant contains a C-terminal deletion of 16 amino acids. The deletion occurred from residue 248 through 264, the end of the wildtype sequence. The resulting protein expresses without the C-terminal UAP56-binding motif.

Human ALYREF-ΔRRM

This is an RNA Recognition Motif (RRM) truncation mutant, truncated from amino acid 114 through 187, resulting in a 73 total residue truncation.

Human ALYREF-ΔNC

Similar to the ALYREF-ΔC16 mutant, this mutant contains the 16 residue C-terminal deletion, as well as an 18 residue deletion of the N-terminus, removing residues 1 through 18.

Human ALYREF-10E

This is the full length protein with ten point mutations to change the positively charged arginine residues found in RGG/RG motifs throughout the wildtype protein to negatively

charged glutamic acid. The mutations selected were: R27E, R30E, R34E, R36E, R38E, R224E, R227E, R231E, R233E, R235E.

Human ALYREF-17E

This mutant is also the full length protein containing the previously described ten point mutations of ALYREF-10E with an additional seven point mutations. The additional mutations were selected because of their positive charge at physiological pH, and their location within the RRM. The total selected mutations were: R27E, R30E, R34E, R36E, R38E, K114E, K140E, K141E, R148E, H159E, R162E, K163E, R224E, R227E, R231E, R233E, R235E.

Cell Transformations

Escherichia coli DH5α competent cells were used for cloning. Cells and DNA were thawed on ice for 30 minutes and incubated together in a 1.5 mL microcentrifuge tube for 15-30 mins, heat shocked in a water bath for 90 s at 37°C, incubated for two min on ice and recovered with 200 µL of autoclaved 1xLB at 37°C for 45 min with shaking. The reaction was diluted 1:10 with 1xLB, and 100 µL of the recovered cells were plated onto an LB-agar plate with the appropriate antibiotic and incubated at 37°C overnight. A single colony was selected from the plate and a suspension culture grown with 3 mL of LB with the appropriate antibiotic. The plasmid DNA was harvested with the QIAprep Spin Miniprep Kit according to the suggested protocol, and the DNA eluted in 40 µL autoclaved sterile ddH_2O.

Rosetta™ competent cells were used for protein expression. Cells and DNA were thawed on ice for 30 min and incubated together in a 1.5 mL microcentrifuge tube for 15-30 min, heat shocked in a water bath for 30-45 s at 42°C, incubated for two min on ice

and recovered with 200 µL of autoclaved 1xLB at 37°C for 45 min with shaking. 100 µL of the recovered cells were plated onto an LB-agar plate with the appropriate antibiotics and incubated at 37°C overnight.

Protein Expression

A small-scale overnight culture was grown by inoculating 2-3 colonies into autoclaved LB with antibiotics, and grown 12-16 h with shaking at 37°C. For each largescale protein expression, 20-25 mL of the small-scale overnight culture were added to each liter of prepared autoclaved LB with antibiotics. The cultures were grown to an OD_{600} between 0.7-0.9 at 37°C with shaking at 160 RPM. Once the proper density was reached, protein expression was induced with 0.2 mM IPTG and incubated at 16°C for 12-16 h. After this time, the cells were spun down to pellets in the Thermo Scientific Sorvall LYNX 6000 Superspeed centrifuge at 13500 RPM for 15 min at 4°C. The pellets were collected, combined and stored in a 50 mL polypropylene tube.

Table 2. Buffers used in purifications. The order shown does not necessarily reflect the order in which the columns were used.

Protein	Lysis Buffer	SP Seph. Wash Buffer	SP Seph. High Salt Buffer	Nickel/His Binding Buffer	Nickel/His Elution Buffer	Mono S Binding Buffer	Mono S Elution Buffer	SEC Buffer	Storage Buffer
ALYREF-WT	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10 % Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10 % Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 300 mM Imid. 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 OR 300 mM KC 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 OR 300 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol
ALYREF-ΔC16	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 500 mM Imid. 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	N/A	N/A	N/A	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol
ALYREF-ΔRRM	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 500 mM Imid. 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	N/A	N/A	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol
ALYREF-ΔNC	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 500 mM Imid. 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	N/A	N/A	N/A	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol
ALYREF-10E	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	150 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	1000 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 500 mM Imid. 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	N/A	N/A	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol	500 mM KCl 20 mM KH_2PO_4 [pH 7.5] 0.05% igepal 1 mM DTT .5 mM EDTA 10% Glycerol

Breaking Cells for Purification

The frozen cell pellet was added to a beaker with 100 mL of lysis buffer (Table 2). Two Pierce protease inhibitor (PI) tablets (Thermo Scientific) or a cocktail of protease inhibitors (aprotinin, chymostatin, leupeptin, and pepstatin A at 5 µg/mL each), and 1 mM phenylmethylsulfonyl fluoride (PMSF) were added. The cells were then mixed at 4°C for 20 min or until all clumps were dissolved. On ice, the cell suspension was sonicated at 50 A in 30 s intervals five times with one-minute rests in between sonication periods. The cell lysate was then divided into 50 mL conical tubes and centrifuged at 13500 RPM at 4°C for 45 min. The supernatant was then separated from the pellet, to prepare for the first column purification. Both supernatant and pellet samples were collected and analyzed by SDS-PAGE and Coomassie Blue staining.

SP Sepharose Chromatography

Approximately 8-10 mL of SP Sepharose® Fast Flow ion exchange beads were loaded to a 50 mL plastic Bio-Rad column, allowed to drain, and equilibrated with 100 mL of wash buffer. The supernatant, flow through, or pooled fractions from a previous column were loaded onto the column with a FPLC pump after being diluted to the same salt concentration as the equilibrated beads, if necessary. The pump was flushed with more lysis buffer, and the flow through from the column was sampled. The column was then attached to the AKTA Pure system for HPLC. The resin was washed with five column volumes of wash buffer. A 100 mL linear gradient from 150 to 850 mM KCl in the buffer was applied to elute ALYREF, and the peak fractions were pooled. The fractions collected were sampled and analyzed by SDS-PAGE and Coomassie Blue staining, and pooled based on band intensity and purity.

Nickel Affinity Chromatography

Approximately 4-5 mL of Nickel High Performance resin was loaded to a 50 mL plastic Bio-Rad column, allowed to drain, and equilibrated with 100 mL of binding buffer. The supernatant, flow through, or pooled fractions from a previous column were loaded onto the column manually or with a FPLC pump, and flushed from the pump with more binding buffer. After loading, the column could be eluted manually, or using the AKTA Pure HPLC system. If eluting manually, the column was washed first with 100 mL binding buffer, then by 20 mL of the same buffer with 5 mM magnesium chloride. Subsequently, another 50 mL of binding buffer was used, followed by a final wash with 30 mL of the binding buffer with 20 mM imidazole. The protein was then eluted with three 4 mL portions of nickel elution buffer (Table 2). The first elution fraction was incubated with the resin for ten min, the second elution fraction incubated for five min, and the third elution with no incubation period.

If eluting by HPLC, the column was attached to the AKTA Pure system immediately after loading the protein solution to the column and not manually washed. The resin was washed by the AKTA Pure system with five column volumes of binding buffer, and the protein was eluted with a 0-85% gradient of elution buffer over a 30column volume span. After eluting, regardless of the method, the fractions collected were sampled and analyzed by SDS-PAGE and Coomassie Blue staining, and pooled based on band intensity and purity.

HisTrap™ Chromatography

The 5 mL HisTrap™ FF column (GE) was prepared according to GE recommended protocol; the column was washed with five column volumes of binding

buffer, five column volumes of elution buffer, and then equilibrated with ten column volumes of binding buffer. The previously pooled fractions were loaded onto a 50 mL GE Superloop, and the sample was applied to the column using the AKTA Pure HPLC system. The column was washed with ten column volumes of binding buffer, and eluted with a 0-85% gradient of elution buffer over 30 column volumes. The wash fraction and eluting fractions were sampled and analyzed by SDS-PAGE and Coomassie Blue staining, and pooled based on band intensity and purity.

Mono S Chromatography

The Mono S cation exchange column was prepared by equilibrating with five column volumes of wash buffer. The previously pooled fractions were diluted to the same salt concentration as the wash buffer and loaded onto a 50 mL GE Superloop, and applied to the column using the AKTA Pure HPLC system. The column was washed with five column volumes of wash buffer, and the protein was eluted with a 0-85% gradient of elution buffer over a 20 column volume span. After eluting, the wash and fractions were sampled and analyzed by SDS-PAGE and Coomassie Blue staining, and pooled based on band intensity and purity. The collected fractions were prepared for size exclusion chromatography by diluting to the same salt concentration as the chosen storage buffer.

Size Exclusion Chromatography

The previously collected fractions were concentrated to at least 3 mL using a Corning Spin-X UF centrifugal concentrator (MWCO 10,000 or 5,000 Da) by centrifugation at 3900 RPM, at 4°C for 15-min increments, checking between cycles for any sign of aggregation. The Superdex 75 Increase 10/300 column was equilibrated with storage buffer, and the concentrated fraction was loaded onto the column in either 250 µL

or 500 µL portions by the AKTA Pure HPLC system. The column was eluted with storage buffer at a rate of 0.2 mL per minute. The chromatogram of A_{280} produced by UNICORN software during the elution proved to not be consistently accurate at determining when ALYREF eluted, so every other or every two fractions collected were sampled and analyzed by SDS-PAGE and Coomassie Blue staining. After analysis, the pooled fractions were then concentrated, aliquoted into small portions, flash-frozen in liquid nitrogen, and stored at -80°C.

Anti-His Western Blot

Anti-His Western blots were used to confirm the presence of the desired protein during purification. A 12% acrylamide SDS gel was run with 2.5 µL of stained ladder, and transferred onto a PVDF membrane prepared with ethanol and incubated in transfer buffer for 2-3 min. The sandwich was assembled using Bio-Rad mini transfer stacks. Bio Rad Trans Blot Turbo transfer system's Mixed Molecular Weight program was used for the transfer (1.3 A, 25 V) with an altered transfer time of 15 min. After transferring, the membrane was rinsed with water and blocked in TBS + 1% casein for one hour at room temperature. After removing blocking buffer, the membrane was incubated for one hour at room temperature in mouse anti-His_6 HRP conjugated antibody at 1:3000 concentration by volume in blocking buffer. Subsequently, the membrane was washed five times with PBS-T in 5-10 min increments, discarding the wash each time. The final wash was in PBS, twice, for five min each. To develop, the membrane was incubated in a 1:1 by volume solution of luminol and peroxide for three min. The membrane was then imaged using the ChemiDoc imaging system and the "chemiluminescent" setting.

Coomassie Blue Staining

Protein content was detected by Coomassie Blue staining solution containing 0.2% (weight/volume) Coomassie Brilliant Blue R-250 powder, 25% methanol, 12.5% glacial acetic acid, and 60% ddH_2O by volume for 30 min at room temperature with shaking, or microwaved for one minute and developed for 10 min with shaking. Gels were destained in destaining solution comprised of 40% methanol, 10% glacial acetic acid, and 50% ddH_2O by volume. Destaining was accomplished by microwaving for one min and destaining to clarity, or by incubating the gel in destain buffer with shaking to clarity. Kim wipes were used to soak up molecules of Coomassie Blue. Destained gels were imaged using the ChemiDoc imaging system and the "stain gel" setting.

Nickel Affinity Pull Down Assay

Pull down assays were used to assess protein-protein interactions between ALYREF and UAP56. For nickel pulldown, 1.5 µg of ALYREF were incubated with 3 µg UAP56 in 30 µL of K buffer with 100 mM KCl and 15 mM imidazole for 30 min at 4°C. The reaction mixture was incubated with 18 µL Nickel High Performance resin (GE Healthcare Life Sciences) for 1.5 h at 4°C. After washing the resin three times with 200 µL of K buffer with 100 mM KCl and 15 mM imidazole, bound proteins were eluted with 20 µL of 2% SDS. 10% of the supernatant (S) and SDS elution (E) fractions, and 2% of the wash (W) fraction were analyzed by 12% SDS-PAGE and Coomassie Blue staining.

Radiolabeling and Synthesis of Substrate

The half R-loop flap substrate was synthesized by labelling ssDNA, annealing with a complementary ssRNA, and annealing again with a complementary ssDNA (reference Table 3). The ssDNA (XX1), and 10X polynucleotide kinase buffer were

thawed on ice, while the ^{32}P-γ-ATP was thawed at room temperature. In a 1.5 mL microcentrifuge tube, 5 µL of ssDNA (8400 ng total), 7 µL of ^{32}P-γ-ATP, 37 µL of sterile ddH_2O, 6 µL 10X PNK buffer (to make a final concentration of 1X PNK) were mixed, and last 5 µL of T4-polynucleotide kinase were added. This mixture was incubated for one h at 37°C, then at 65°C for 10 min. During this time, two Micro Bio-Spin P-6 columns were prepared by removing the snap off tips and expelling the liquid inside the columns by capping and un-capping the tops. The columns were then centrifuged at 4000 RPM for four min to remove the rest of the liquid. The labelled ssDNA solution was then added to one P-6 column and subjected to centrifugation at 4000 RPM for four min. The flow through was collected and added to the second P-6 column which was spun down again at 4000 RPM for four min. The flow through was collected again and the concentration calculated using the Nanodrop DNA setting.

Using the calculated concentration, the first annealing can be performed. Using the full amount of flow through collected, a short ssRNA section (R-5'F) was added in approximately three times the calculated amount of DNA in ng. 10X buffer H (500 mM Tris [pH 7.5], 100 mM $MgCl_2$, 1 M NaCl, 10 mM DTT) was added to the DNA-RNA solution to make a final concentration of 1X buffer H, and a small amount of sterile ddH_2O was added to assist in bringing the total volume of the solution to a multiple of 10. The solution was then divided into two PCR tubes equally, and run on an annealing program with a lid temperature of 30°C, holding temperature of 95°C followed by a decrease of 0.1°C every 15 s until 25°C was reached and held for five min. Subsequently, the temperature was maintained at 4°C.

The solution from the PCR tubes was recombined in a 1.5 mL microcentrifuge tube. An approximately equal concentration of complementary ssDNA (XX2) was mixed with enough 10X buffer H to yield a final concentration of 1X buffer H. This solution was added to the labelled DNA:RNA substrate and incubated for one h at room temperature to complete the second annealing.

Table 3. Oligos used in this research.

Oligo Name	RNA or DNA	Length	Sequence
XX1	DNA	60	5'ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC3'
XX2	DNA	60	5'GGGTGAACCTGCAGGTGGGCAAAGATGTCCCAGCAAGGCACTGGTAGAATTCGGCAGCGT3'
R-5' F	RNA	30	5'GGGUGAACCUGCAGGUGGGCAAAGAUGUCC3'

Purification of Radiolabeled Substrate

A 7% polyacrylamide native TAE (40 mM Tris base, 20 mM acetic acid and 1 mM EDTA) or TBE (40 mM Tris base, 20 mM boric acid and 1 mM EDTA) gel was made within large glass plates, with spacers, and a large comb for use in the Bio-Rad Protein II xi Cell. Thin strips of Parafilm were stretched on the underside of the plates to prevent leaking while polymerization occurred. Once polymerized, the gel was set up in the Bio-Rad Protein II xi Cell with 3 L of 1X TAE or TBE, using water circulation to maintain the temperature at 4°C. 30 µL of 6X DNA loading dye was added to the substrate, and all of the mix was loaded into the middle lane of the gel. The gel was run at 110V for three h, or until the dye reached about two thirds of the way down the gel. After three h, the gel was removed from the cassette and viewed in a dark room with a shortwave UV lamp. The uppermost band was excised, and added to a dialysis tube with

2.5 mL of 1X TAE or TBE buffer in a horizontal electrophoresis chamber. The chamber was set up with 1X TAE or TBE buffer, and run at 110 V for 50 min. The supernatant from the dialysis tube was collected and set aside in a 14 mL conical tube on ice. Another 2.5 mL of 1X TAE was added to the dialysis tube where the gel slice remained, and the system was run again at 110 V for 50 minutes. The supernatant was collected again and added to the 14 mL conical tube. 5 mM $MgCl_2$ was added to the supernatant to increase stability of the substrate through cationic interactions, and it was then transferred to a Corning Spin-X UF centrifugal concentrator (MWCO 5,000 Da). The substrate was concentrated at 3900 RPM at 4°C in 10 min increments until about 200 µL remained. 2.5 mL of TMN buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl) were added to the concentrator, concentrated again to 200 µL, and repeated once more. The substrate was then aliquoted equally into multiple 1.5 mL microcentrifuge tubes, and stored at -20°C.

R-Loop Resolution Assays

R-loop resolution assays were used to determine the rate at which R-loops are resolved in vitro with the addition of UAP56 and ALYREF at varying concentrations. Assays were performed in 1.5 mL microcentrifuge tubes. UAP56 (with or without ALYREF) was incubated with 5 nM of fresh R-loop substrate in reaction buffer (25 mM Tris-HCl, 1 mM DTT, 1 µg BSA, 2 mM ATP, 1 mM $MgCl_2$) at 30°C for 20 min. It is imperative to use R-loop substrate that has not gone through freeze-thaw cycles, as substrate background greatly increases with the occurrence of these cycles. Reaction mixtures were deproteinized before being resolved in 7% polyacrylamide native TAE or TBE buffer at 4°C and analyzed. The gels were dried onto Amersham Hybond-N+ membrane at 80°C for one hour, developed on a phosphor screen overnight, and imaged

with the PharosFX Plus Molecular Imager system.

Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays were used to determine the extent to which a protein bound to RNA, without the presence of ATP. ALYREF-WT or mutant (12.5 to 100 nM) was incubated with 5 nM R-loop substrate at 30°C for 10 min in 8 µL of reaction buffer (25 mM Tris-HCl, 1 mM DTT, 1 µg BSA, 1 mM $MgCl_2$, 100 mM KCl). The reaction mixtures were resolved in 7% polyacrylamide gels in TAE buffer at 4°C. The gels were dried onto Amersham Hybond-N+ membrane at 80°C for one h, developed on a phosphor screen overnight, and imaged with the PharosFX Plus Molecular Imager system.

III. RESULTS AND DISCUSSION

ALYREF-WT expression and purification

Rosetta™ competent cells were transformed with ALYREF-WT-pET24b and grown to mid-log phase, taking samples at various timepoints to time induction between an OD_{600} of 0.6 and 0.8. Protein expression was induced by 0.2 mM of IPTG. Rosetta cell pellets expressing ALYREF-WT were lysed by sonication in 150 mM K buffer with 0.05% Igepal, and initially subjected to purification by an SP Sepharose® Fast Flow chromatography column with increasing salt concentration. The fractions were analyzed by gel electrophoresis (Figure 4). The elution fractions showed a strong band between 25 and 37 kDa, which was presumed to be ALYREF-WT since its predicted molecular weight is 27.6 kDa. Elution fractions in lanes 11 through 23 showed the most concentrated protein with the least amount of lower molecular weight products. These fractions were selected, pooled and diluted to 500 mM KCl concentration for nickel purification.

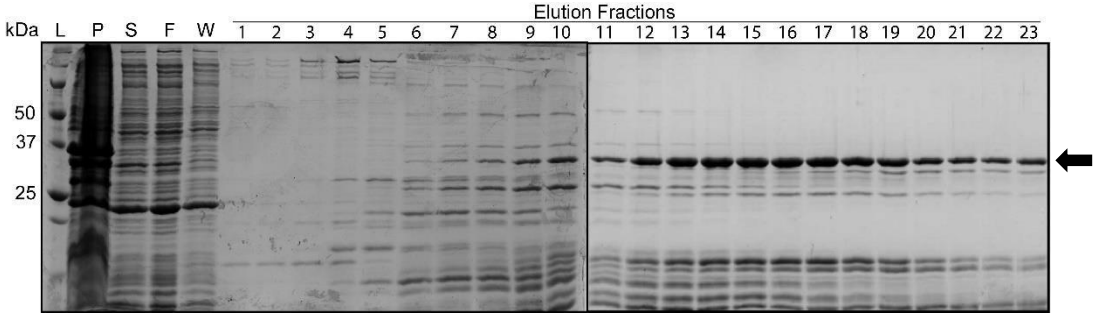

Figure 4. ALYREF-WT affinity chromatography fractions. The cell lysate was purified using SP Sepharose® Fast Flow chromatography ("L", ladder, "P", pellet, "S", supernatant, "F", flow through, "W", wash). The fractions were separated by gel electrophoresis and stained with Coomassie Blue. In the elution, the dark bands between 25 and 37 kDa are ALYREF-WT.

With a His_6 tag on the C-terminus, ALYREF-WT was expected to have a high affinity for nickel resin. The fractions from the SP Sepharose column were loaded onto a nickel affinity column, and washed with 500 mM KCl buffer, and subsequently with low salt buffer with 5 mM magnesium chloride to prevent non-specific binding. Another wash with low-salt buffer with 20 mM imidazole was done to elute any weakly bound protein. To elute ALYREF-WT, three successive elutions of 500 mM KCl buffer with 300 mM imidazole were performed, and a sample of the nickel beads after elution were also taken to determine if any protein was still bound to the column (Figure 5). All three elutions, between 25 and 37 kDa, appear to contain a large amount of ALYREF-WT, so they were combined to make a total of 12 mL. Since the elutions were in high-imidazole buffer, the fractions were diluted to a salt concentration of approximately 150 mM to proceed with the next step in purification.

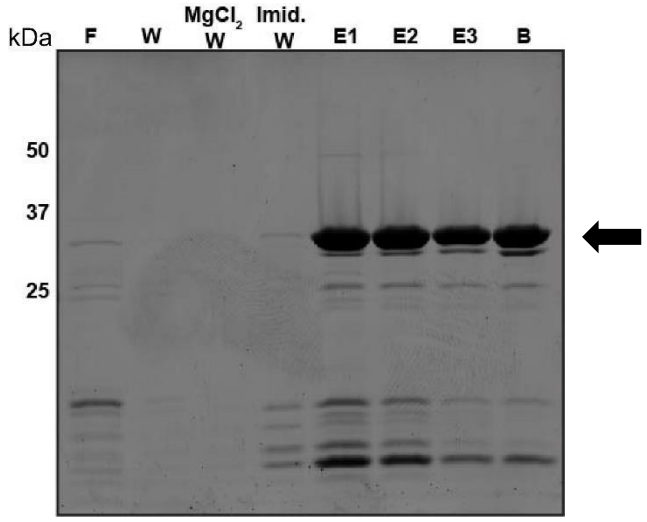

Figure 5. ALYREF-WT nickel affinity chromatography fractions. The previously pooled fractions were further purified with nickel affinity chromatography ("F", flow-through, "W", wash, "E", elution, "B", beads). The fractions were separated by gel electrophoresis and stained with Coomassie Blue.

The Mono S column was used to further purify ALYREF-WT by utilizing cation exchange and prepare for size exclusion chromatography. The protein was eluted with increasingly higher salt concentration, and resulted in fractions 1-11 being collected and pooled (Figure 6). The fractions were concentrated to approximately 1.6 mL, and run in 500 µL amounts for each run of gel filtration using the Superdex® 75 Increase with K+ 300 mM KCl + 0.01% igepal +1 mM DTT as the running buffer. Despite concentrating the cation exchange fractions from 12 mL to 1.5 mL in an effort to decrease the total amount of elution to run on the size exclusion column, the SDS-PAGE analysis of the fractions shows a much lower amount of protein than expected. This could be attributed to leakage while concentrating, or aggregation of protein. ALYREF-WT was found to have a sharp, significant UV peak at 9.9 mL using absorbance 280 nm, indicating that the purified ALYREF protein behaved well and was not aggregated, since aggregated proteins typically have broad peaks. However, the UV peak was found to not be an accurate marker for where ALYREF elutes, and instead it was discovered in fractions immediately after the peak at about 10.2 mL. This is unsurprising, since the molar extinction efficient was calculated to be 11460 $M^{-1}cm^{-1}$, indicating that the absorbance of a 1 mg/mL solution at 280nm is 0.416. K_{av} values and elution volume data from a series of standards run on this column allowed linear regression analysis to be performed, which yielded this equation to calculate the experimental molecular weight:

$$y = -.3725x + 1.9102$$

The K_{av} is represented by y and x is the $\log_{10}$ of the molecular weight of the standard (Da). Using this equation, the experimental molecular weight was calculated to be 54.92 kDa, which indicates ALYREF-WT is likely eluting as a dimer. Fractions 7 through 10

were collected, pooled, concentrated with a 10 kDa concentrator, aliquoted into 5 µL portions, and stored at -80°C.

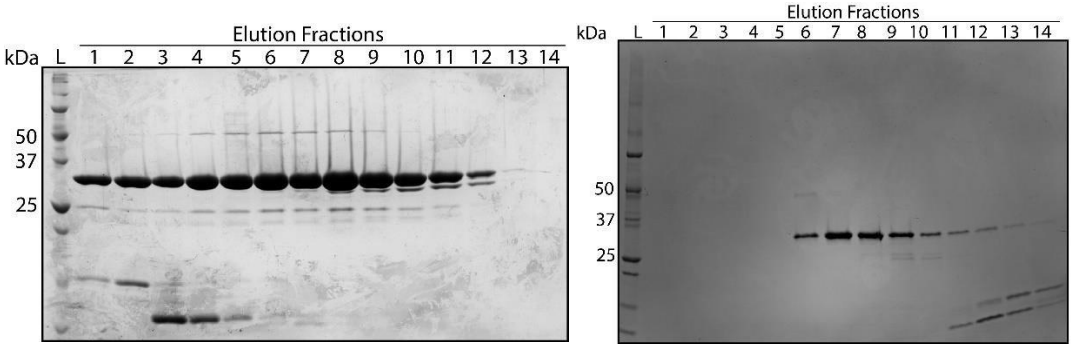

Figure 6. Mono S and Superdex 75 elution fractions comparison. All of the Mono S fractions except for the last were pooled and concentrated. Shown in this figure is one of the gels with samples from the size exclusion chromatography fractions alongside the gel with cation exchange fractions.

ALYREF-WT and UAP56 pull down assay

To prove interaction between ALYREF through its previously described C-terminal UAP56 binding motif and UAP56-WT and helicase dead mutants, pull-down assays were performed using nickel resin which ALYREF binds to. Purified ALYREF and UAP56 were combined to incubate together in 100 mM KCl solution for 30 minutes. ALYREF and UAP56 are shown in the gel, respectively, at around 30 kDa and 50 kDa (Figure 7). The purpose of this assay is to demonstrate that ALYREF, a known binder of nickel resin, will "pull down" UAP56 from solution since the proteins associate with each other. The supernatant ("S") of each assay contained excess UAP56, and excess ALYREF that did not bind to the small amount of nickel resin used. The wash samples ("W") were taken after the fourth wash of the resin, and none showed any protein, meaning that the ALYREF did have a high affinity for the nickel resin and could not be knocked-off with low-salt low-imidazole washes. The elutions ("E") consisted of 8 µL samples of the nickel resin combined with 20 µL of 2X SDS loading dye. Each elution,

aside from the controls, showed approximately equal amounts of ALYREF and UAP56-WT and mutants, as expected, indicating that there is interaction between the two proteins. Furthermore, since the elutions from both mutant assays show interaction between the two proteins, this indicates that the ability of UAP56 to associate with ALYREF is not dependent on its ATP-binding or hydrolysis capabilities. Each control assay containing only UAP56 showed no protein in the wash or elution fractions, indicating that neither UAP56-WT nor its mutants associate with nickel resin, and remain entirely in the supernatant.

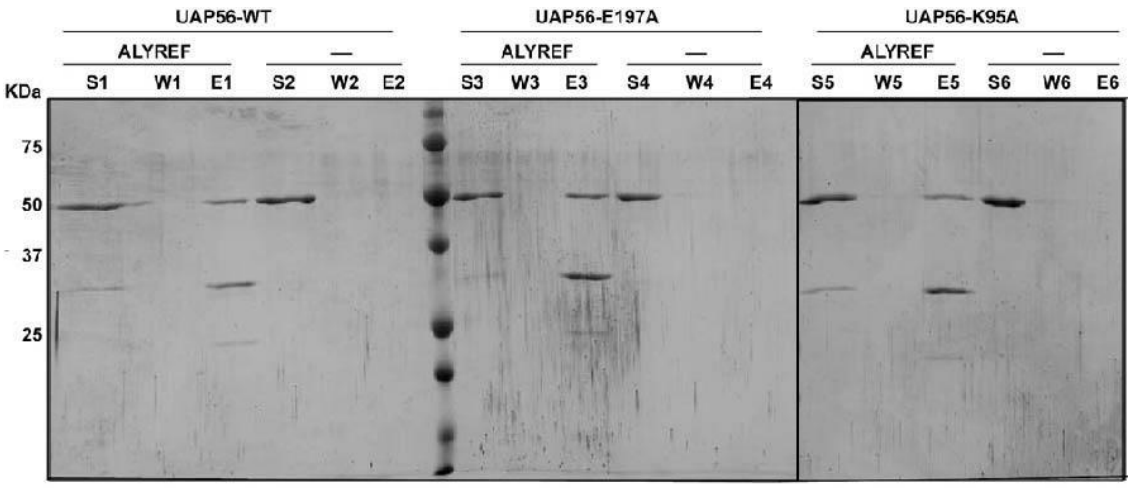

Figure 7. Pull down assay with ALYREF-WT, UAP56-WT and its two mutants K95A, and E197A. The first three lanes of each assay contain both ALYREF and UAP56, with the following three lanes used as controls with no ALYREF ("S", supernatant, "W", wash, "E" elution). Each assay and its control are distinctly indicated by the UAP56 label separating them.

Initial testing of the R-loop resolution capabilities of UAP56

After confirming the interaction between UAP56 and ALYREF exists, 5' RNA:DNA flap structure that resembles a branch migratable R-loop structure was constructed. A radioactive phosphate label was initially added to the 5' end of ssDNA before annealing with the ssRNA 30mer, followed by a second annealing with a

complementary ssDNA. The annealing of these three strands created a “flap” structure to be used to simulate R-loops that may occur *in vivo* from the binding of nascent mRNA to its parent template. This R-loop structure was purified, then standardized to 50 nM for use in assays, making the final concentration 5 nM. The quantification was done by subtracting the background in lane 1 from the other lanes, establishing the background as the baseline of zero dissociation (Figure 8A). The top band in each assay is representative of the structures that have not been resolved, and still contain the complementary RNA strand. The lower bands are representative of resolved structures without the RNA 30mer, thus making the structures less bulky in size and mass, allowing them to travel further in the native gel. The radioisotope label remains, and allows quantification. The average resolution of the substrate at a UAP56-WT concentration of 200 nM was determined to be 80.8 ± 1.55% (Figure 8B). The wildtype’s activity appears to decrease significantly at concentrations less than 100 nM. The UAP56-K95A mutant lacks the ability to bind to ATP, which prevents helicase activity since UAP56 is an ATP-dependent RNA helicase. This mutant’s R-loop resolution activity was quantifiably negligible, which is consistent with the hypothesis that the K95A mutation prevents helicase activity.

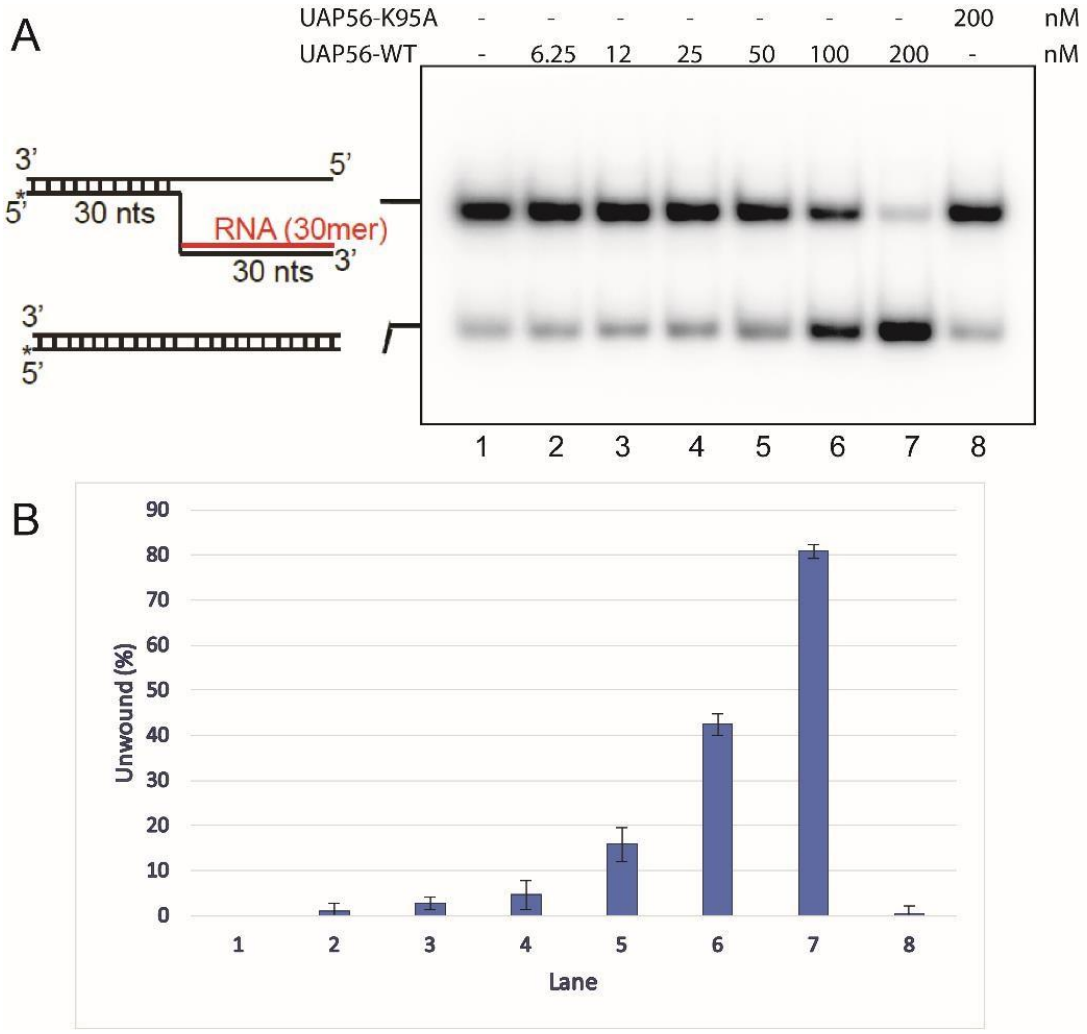

Figure 8. UAP56 unwinds DNA:RNA flap structures mimicking R-loops.
A) DNA:RNA unwinding assay with UAP56-WT and UAP56-K95A using DNA:RNA flap structures mimicking R-loops as substrates. The first lane shows the substrate alone. Lanes 2 through 7 show the resolution of R-loops by UAP56-WT (6.25-200 nM). Lane 8 shows that resolution by the helicase dead mutant K95A at a 200 nM concentration. B) Graphical representation of the mean of the percentage of dsDNA product recovered after the reaction, with error bars showing standard deviation in percent (n=3).

Testing of the R-loop resolution capabilities of UAP56 with ALYREF-WT

Based on the results from previous preliminary assays, it was decided that low concentrations of ALYREF should be used to determine if there is a threshold at which resolution of R-loops increases greatly. The next assay performed was a "time course" assay to observe the correlation between reaction time and percent dissociation of the

substrate (Figure 9). The assay was performed in triplicate, using a 12.5 nM concentration of ALYREF, and 25 nM concentration of UAP56. The substrate used for these assays was newly synthesized, which yielded a lower background and the ability to more accurately quantify the data. The first lane was used as the control, containing no UAP56 or ALYREF (Figure 9A). Lanes 2 through 4 show the amount unwound by only UAP56 at time intervals of 5, 10, and 15 minutes. At 15 minutes, the percent dissociation by UAP56 was 8.14 ± 4.79% (Figure 9B). For ALYREF at 15 min, the percent dissociation was 25.48 ± 1.56%. Using both ALYREF and UAP56 to resolve the R-loops resulted in a percent dissociation of 84.6% ± 0.783% after 15 minutes (Figure 9B, lane 7). Using the lower concentration of ALYREF, the results became much clearer that ALYREF does appear to regulate the DNA/RNA helicase and R-loop dissociation activities of UAP56. The average dissociation of R-loops by 25 nM UAP56 at 15 min, added with the average dissociation of R-loops by 12.5 nM of ALYREF at 15 min yields 33.6%. The percent dissociation of the combined proteins far exceeds the dissociation of the individual proteins, which indicates ALYREF increases the R-loop dissociation activity of UAP56.

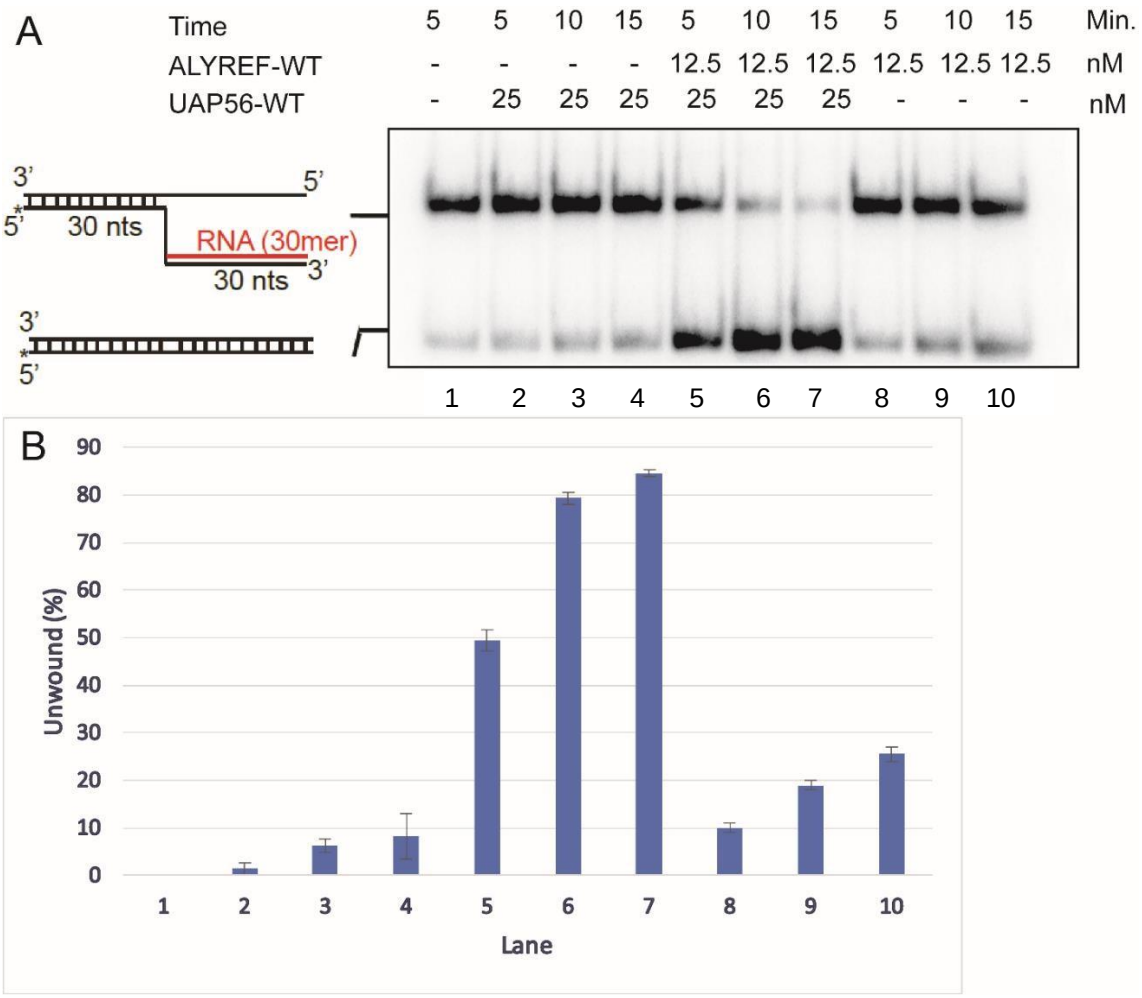

Figure 9. ALYREF-WT stimulates the R-loop resolution activity of UAP56 significantly. A) A time course assay comparing the R-loop resolution by UAP56 alone, ALYREF alone, and UAP56 and ALYREF together. The first lane shows the substrate alone. Lanes 2-4 show the resolution of R-loops by 25 nM UAP56 for 5-15 minutes. Lanes 5-7 show the resolution of R-loops by both 12.5 nM ALYREF and 25 nM UAP56 for 5-15 minutes, and lanes 8-10 show resolution of R-loops by 12.5 nM ALYREF alone for 5-15 minutes. B) Graphical representation of the mean of the percentage of dsDNA product recovered after the reaction, with error bars showing standard deviation in percent (n=3).

After collecting data that suggested ALYREF upregulates the R-loop dissociation activity of UAP56, it was decided that an even lower concentration of ALYREF should be tested to determine if that regulation could occur with even less protein. 25 nM and 50 nM UAP56 were used alone in two lanes, and in combination with 6.25 and 12.5 nM

ALYREF, which was also used independently in two lanes as well (Figure 10A). This assay was performed in triplicate. Containing no ALYREF or UAP56, lane 1 served as the control. Lane 2 contained 25 nM UAP56, and resulted in 0.645 ± 1.12% dissociation (Figure 10B). Lane 3 contained 50 nM UAP56 and resulted in 16.1 ± 2.88% dissociation. Lanes 4 through 6 range from 63.5 ± 20.8% to 77.1 ± 7.30%. These values are not statistically different enough to make any specific assumptions about the cumulative effect of UAP56 and ALYREF at these concentrations. However, the general trend of the R-loop dissociation ability of UAP56 increasing multiplicatively with the addition of ALYREF remains consistent throughout these assays. In lane 7, containing 50 nM UAP56 and 12.5 nM ALYREF the highest dissociation was observed at 88.3 ± 2.90%, and there is an observable difference on the gel in comparison to lane 6. Lane 8 contained 6.25 nM ALYREF and yielded a dissociation of 8.21 ± 3.60%, which is similar to 12.5 nM ALYREF in lane 9 which resulted in dissociation of 10.3 ± 4.15% of the substrate. Once again, the additive percent dissociations of ALYREF and UAP56 independently are significantly less than the cumulative effect both have in solution together. Taken together, these experiments support the hypothesis that ALYREF could significantly stimulate the R-loop resolution activity of UAP56.

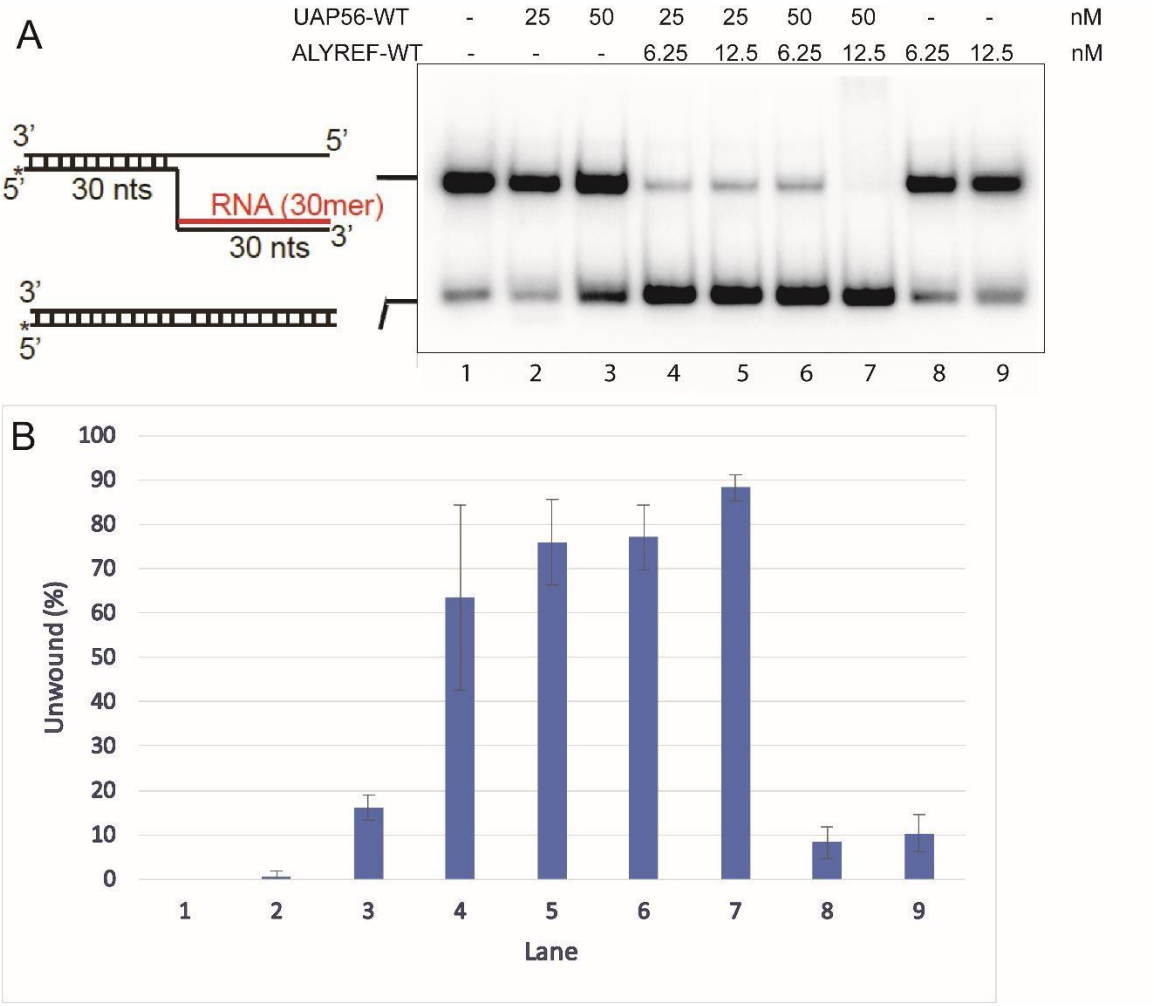

Figure 10. ALYREF stimulates the R-loop resolution activity by UAP56 significantly. A) The titration assay showed that ALYREF-WT stimulates UAP56's R-loop resolution activity. The first lane shows the substrate alone. 25 nM and 50 nM UAP56 were used, as well as 6.25 nM and 12.5 nM ALYREF, as indicated. B) Graphical representation of the mean of the percentage of dsDNA product recovered after the reaction, with error bars showing standard deviation in percent (n=3).

ALYREF-ΔRRM and ΔC16 expression

The data suggesting that ALYREF does significantly increase the ability of UAP56-WT to resolve R-loops prompted the exploration of the ability of ALYREF to bind to UAP56, and if the interaction is required for the stimulation of R-loop resolution.

For this purpose, ALYREF-ΔC16 was designed, which removes the 16 C-terminal amino acid residues that are known to associate with UAP56 [27]. ALYREF also has a conserved RNA recognition motif that lies between residues 114 and 187.

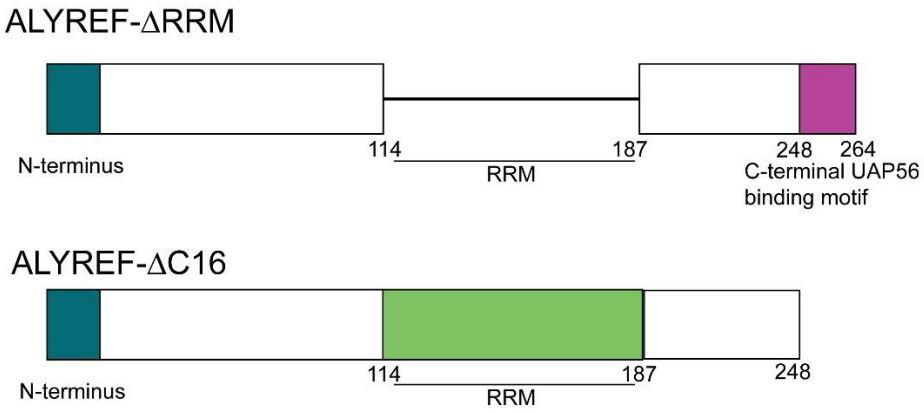

Figure 11. Schematic representation of ALYREF-ΔRRM and ΔC16. Relevant domain structure of ALYREF depicting the of the RRM in lime green, the N-terminus in teal, and the C-terminal UAP56 binding motif in magenta.

A mutant was constructed with this motif in mind, removing that section of the protein entirely, which would theoretically prevent it from binding to RNA, thus reducing or eliminating its effect on UAP56 (Figure 11). These mutant plasmids were synthesized by Gene Universal in pET24a vectors. Both mutants were received and run on a 0.7% agarose gel, which revealed that the plasmids were mainly in nicked open-circular form, not the supercoiled form that would be needed for transformation into Rosetta cells for protein expression. Both plasmids were then transformed into DH5α, and grown on kanamycin plates overnight at 37°C. Colonies were selected from the plates and suspended in 3 mL of LB containing 50 μg/mL kanamycin and grown overnight at 37°C with shaking. The following day, the DNA was harvested from the cells using Qiagen's spin miniprep kit. 1 μL of each mutant plasmid was sampled and run on a 0.7% agarose gel which revealed the majority of the DNA was in the

supercoiled state (Figure 12). These new mutants were compared to ALYREF-WT to confirm that they are both the correct size.

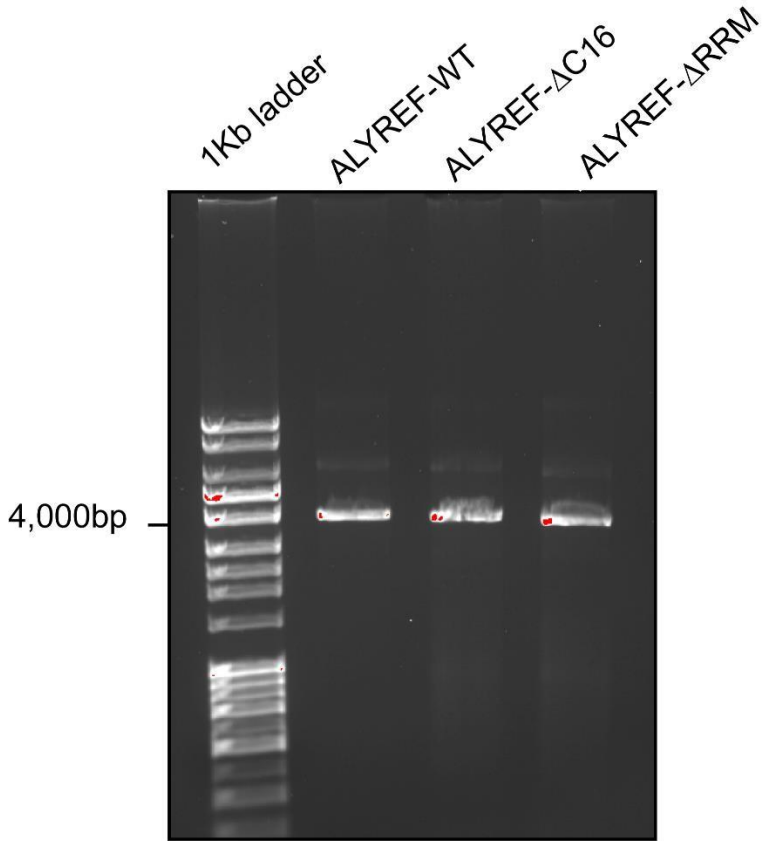

Figure 12. Agarose gel comparing ALYREF-WT and two mutants. The brightest visible band shows supercoiled DNA.

ALYREF-ΔRRM purification

After confirming the plasmids were correct, transformations into Rosetta cells were performed. Each transformation was plated with kanamycin and chloramphenicol, and incubated overnight at 37°C. The following day, a small-scale culture (75 mL) of each was started with antibiotics, and grown overnight at 37°C with shaking. The cell cultures were distributed into three flasks of 1 L portions of autoclaved LB with antibiotics for each mutant. The flasks were incubated at 37°C with shaking until OD_{600} reached 0.7-0.9. Then, protein expression was induced with 0.2 mM IPTG and incubated

at 16°C overnight with shaking. The cells were then harvested the next day, and stored at -80°C.

The cell pellet expressing ALYREF-ΔRRM were lysed by sonication 100 mL of K + 150 mM KCl + 0.05% igepal + 1 mM DTT. The first column used to initially purify the protein was an SP Sepharose chromatography column with increasing salt concentration. The fractions were analyzed by gel electrophoresis. The second gel contained the majority of the eluted ALYREF, and fractions corresponding to lanes 15 through 19 were collected and combined (Figure 13).

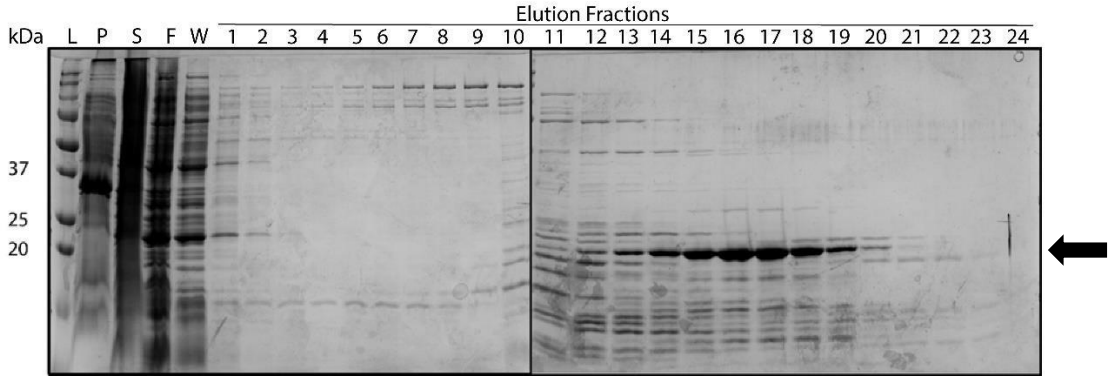

Figure 13. ALYREF-ΔRRM affinity chromatography fractions.
The cell lysate was purified using SP-Sepharose® Fast Flow affinity chromatography ("L", ladder, "P", pellet, "S", supernatant, "F", flow-through, "W", wash). The fractions were separated by gel electrophoresis and stained with Coomassie Blue. ALYREF-ΔRRM elutes between 20 and 25 kDa, while its molecular weight is presumed to be 19.4 kDa.

After the fractions were combined, the HisTrap™ High Performance column was used, taking advantage of the histidine tag. 16.8 mL of elution from the SP Sepharose column was loaded onto the column by Superloop. The binding buffer used was K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and the elution buffer was the binding buffer with 500 mM imidazole added. This column worked very well to further separate the protein from contaminants, and resulted in the last four fractions from the first gel and

the first fraction from the second gel to be collected and combined (Figure 14). The fractions were then combined, and concentrated in a 5 kDa MWCO centrifugal concentrator. The final volume after concentrating was 4 mL. 3 mL of the concentrated protein was set aside, and flash frozen in liquid nitrogen for later use.

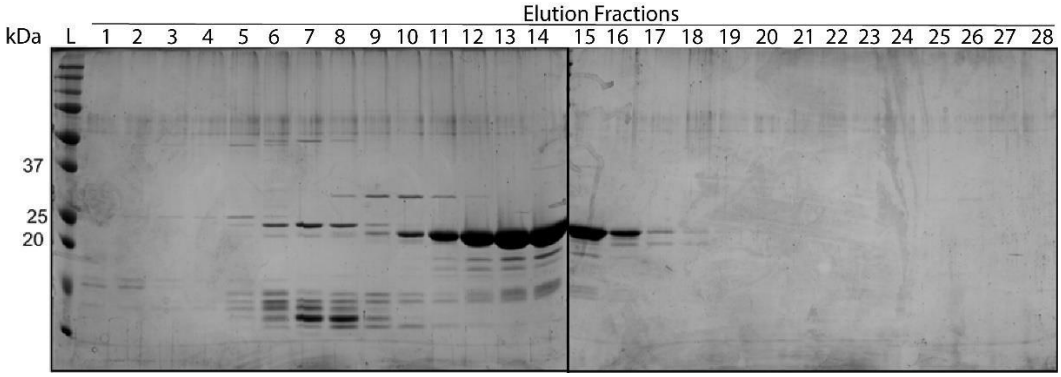

Figure 14. ALYREF-ΔRRM cation exchange chromatography fractions. ALYREF-ΔRRM eluted using the 1 mL HisTrap™ High Performance column. The five most concentrated fractions were selected and pooled for further purification.

The remaining 1 mL was used to further purify by size exclusion chromatography, using the Superdex 75 Increase. 500 µL at a time were loaded to the column, and eluted with K + 500 mM KCl + 0.05% igepal + 1 mM DTT. The slight peak from absorbance 280 nm still remained an unreliable indicator of when ALYREF-ΔRRM eluted, so every other elution fraction was sampled and analyzed by SDS-PAGE and Coomassie staining (Figure 15). The three most concentrated and pure fractions were selected from each run of size exclusion chromatography and combined. The elution was concentrated by a 5 kDa MWCO centrifugal concentrator from 3 mL to 1 mL, and aliquoted into 5 µL portions. The aliquots were flash frozen in liquid nitrogen, and stored at -80°C.

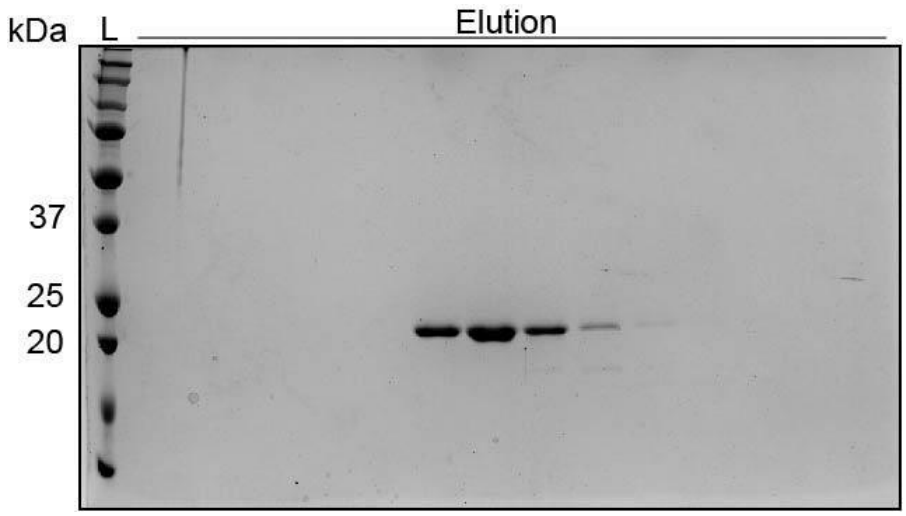

Figure 15. ALYREF-ΔRRM size exclusion chromatography fractions. Size exclusion chromatography was performed twice with 500 µL portions for each run. The samples from each elution resulted in two identical SDS-PAGE, stained with Coomassie Blue. The three most concentrated and pure fractions were selected to be concentrated and aliquoted.

ALYREF-ΔC16 purification

Purification for the second mutant began with the breaking of cells in K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and sonication. The cell lysate was separated by centrifugation into supernatant and pellet, and sampled. The pellet appeared to contain a large amount of the ALYREF protein still. The supernatant was diluted to a salt concentration of 150 mM. The first column used to initially purify the protein was an SP Sepharose column. After the supernatant was loaded onto the column, the protein was eluted by HPLC, and sampled. The second gel contained the majority of the eluted ALYREF, and fractions 15 through 21 were collected and combined to give a total of 25.5 mL (Figure 16). ALYREF-ΔC16 has a presumed molecular weight of 25.8 kDa.

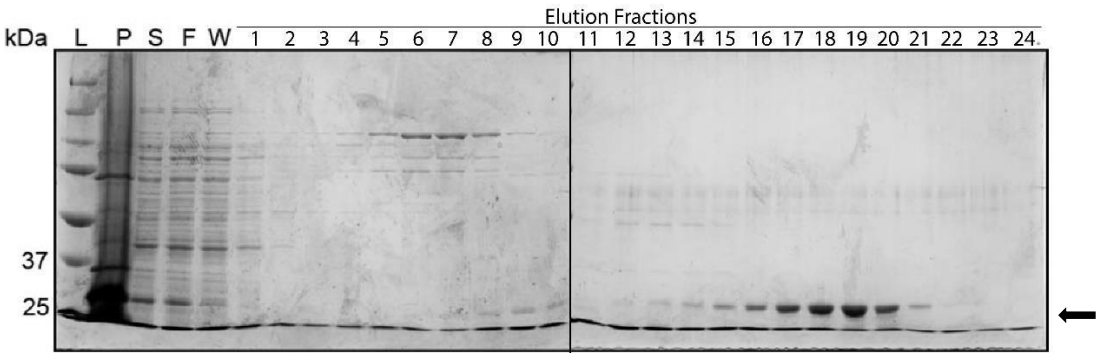

Figure 16. ALYREF-ΔC16 affinity chromatography fractions. The cell lysate was purified using SP-Sepharose® Fast Flow chromatography ("L", ladder, "P", pellet, "S", supernatant, "F", flow-through, "W", wash). The fractions were separated by gel electrophoresis and stained with Coomassie Blue. ALYREF-ΔC16 elutes between 25 and 37 kDa.

The second column used to further purify ALYREF-ΔC16 was the HisTrap™. 25.5 mL of elution from the SP Sepharose column was loaded to the column by Superloop. The binding buffer used was K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and the elution buffer was the binding buffer with 500 mM imidazole added. The protein was eluted with a linear gradient of 0-85% elution buffer. The column worked extremely well, and fractions 16 through 20 were collected since the SDS-PAGE analysis revealed they were the most pure and concentrated (Figure 17). It was decided that the two columns were sufficient for the purification of this mutant, so the combined fractions were prepared for storage. Half of the fraction mixture was flash frozen in liquid nitrogen and stored at -80°C for future use. The other half needed to undergo a buffer change to rid it of imidazole from the high imidazole elution. Concentration was performed using a 5 kDa MWCO centrifugal concentrator, changing the buffer to K + 500 mM KCl + 0.05% igepal + 1 mM DTT twice after concentrating to 500 µL each time. The final amount was 300 µL, which were aliquoted into 10 µL portions, flash frozen in liquid nitrogen, and stored at -80°C.

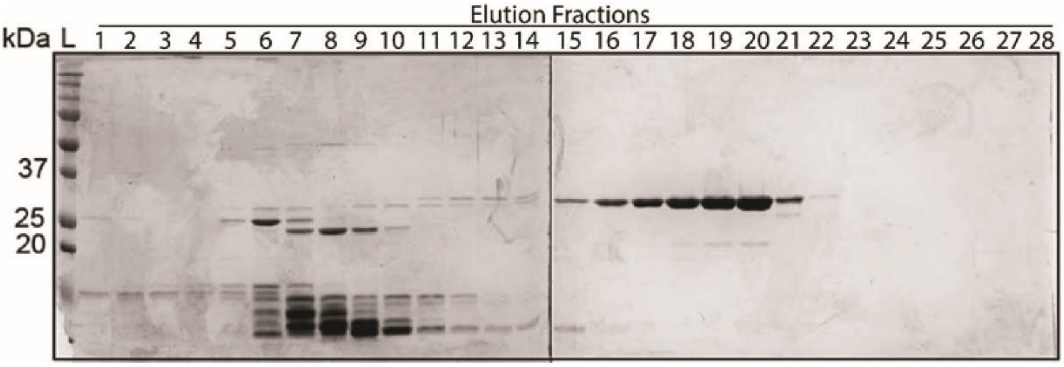

Figure 17. ALYREF-ΔC16 cation exchange chromatography fractions. ALYREF-ΔC16 eluted using the 1 mL HisTrap™ High Performance column. The five most concentrated fractions were selected and pooled for further purification.

Concentration analysis of ALYREF proteins

An SDS-PAGE gel was run with bovine serum albumin standards of 0.5 µg and 1 µg, 0.5 µL and 1 µL of each ALYREF variant (Figure 18). ALYREF-WT and ALYREF-ΔC16 were determined to have concentrations of 300 µg/mL and 400 µg/mL, and the ALYREF-ΔRRM was determined to have a concentration of 1.2 µg/µL. These concentrations were determined in preparation for a pull down assay.

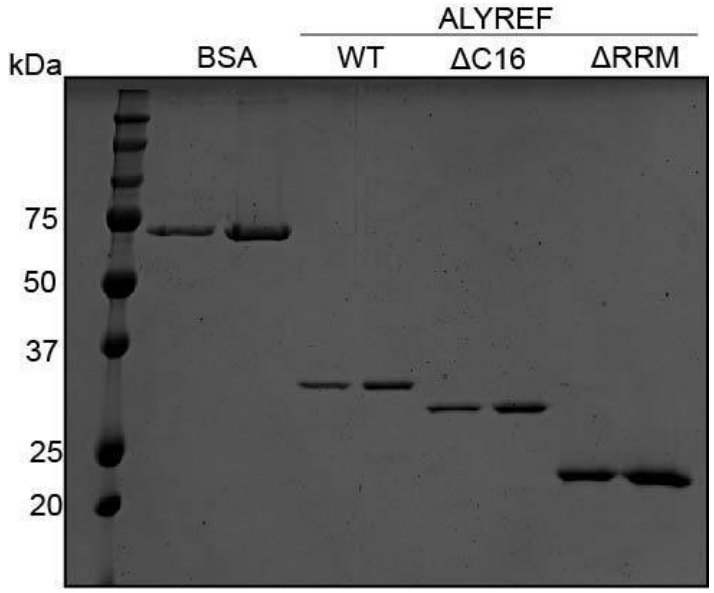

Figure 18. SDS-PAGE gel for concentration analysis. ALYREF-WT, ALYREF-ΔC16, and ALYREF-ΔRRM in shown in comparison to bovine serum albumin standards.

ALYREF-ΔC16, ALYREF-ΔRRM and UAP56 pull down assay

To verify that interaction between ALYREF-ΔC16 and UAP56 was disrupted as hypothesized, and that ALYREF-ΔRRM maintained its interaction with UAP56, pulldown assays were performed using nickel resin which ALYREF binds to. All assays were performed with a final concentration of 100 mM KCl. Purified ALYREF-WT and UAP56 were combined to incubate together and be sampled as a positive control. The second assay was performed without ALYREF, as a negative control, assuming that UAP56 would elute entirely in the supernatant. The supernatant of each assay contained excess UAP56, and excess ALYREF that did not bind to the small amount of nickel resin used (Figure 19, lanes labeled "S"). The wash samples were taken after the fourth wash of the resin, and none showed any protein, meaning that the ALYREF did have a high affinity for the nickel resin and could not be knocked-off with low-salt low-imidazole washes (Figure 19, lanes labeled "W"). The elutions consisted of 8 µL samples of the nickel resin combined with 20 µL of 2X SDS loading dye.

The first assay with ALYREF-WT had results that were consistent with previous findings, indicating that ALYREF-WT and UAP56 associate together, and that UAP56 is able to be pulled-down by ALYREF-WT (Figure 19, lane 4). The negative control was consistent with previous findings, indicating that UAP56 does not interact with the nickel beads. It should also be noted that the UAP56 in this assay appears different than the UAP56 in the previous pull down assay. It was previously discovered during a concentration comparison that different tubes of aliquoted UAP56 contained varying purities of UAP56, indicating that when the protein was concentrated and aliquoted, it was not sufficiently mixed to ensure homogeneity. This was remedied by recombining all

aliquots, mixing, and redistributing. So, the UAP56 in this assay contains a light band just below the molecular weight of ALYREF-WT, which is evident in the negative control assay.

The assay with ALYREF-ΔC16 showed an excess of UAP56 in the supernatant that did not bind with ALYREF, likely because of excess UAP56 present rather than inhibition of interaction between the two proteins as hypothesized. The wash contained no protein, but the elution contained ALYREF-ΔC16 and UAP56 in approximately equal amounts, indicating that mutant could be as effective as the wildtype protein in binding with UAP56 (Figure 19, lane 10). ALYREF-ΔRRM maintained its ability to interact with UAP56, as expected, since both are present in approximately equal amounts in the elution (Figure 19, lane 13). The supernatant also contained excess ALYREF-ΔRRM, indicating there was an excessive amount of both proteins that were unable to be bound to the nickel resin.

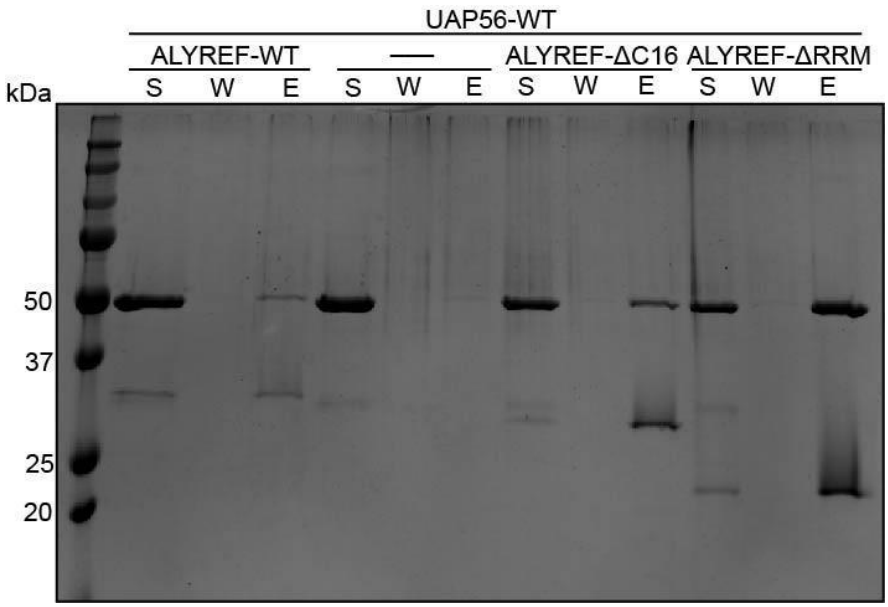

Figure 19. Pull down assay with ALYREF-WT, ALYREF-ΔC16, ALYREF-ΔRRM, and UAP56-WT. The three lanes of each assay contain both ALYREF and UAP56, aside from the control which contains no ALYREF ("S", supernatant, "W", wash, "E" elution). Each assay is distinctly indicated by the ALYREF label separating them.

Testing of the R-loop resolution capabilities of UAP56 with ALYREF-ΔC16 and ALYREF-ΔRRM

After the discovery that ALYREF-ΔC16 maintained its ability to bind to UAP56, the decision was made to proceed with testing its effect on the R-loop resolution activity of UAP56. The lowest concentrations of ALYREF-WT that were originally tested, 6.25 nM and 12.5 nM, were used to test the two mutants, and UAP56 continued to be tested with the standard concentration of 50 nM (Figure 20A). Both mutants appeared to dissociate R-loops by themselves at least twice the rate of ALYREF-WT, which was unexpected. Containing no ALYREF or UAP56, lane 1 served as the control. Lane 2 contained 6.25 nM of ALYREF-ΔC16, and resulted in 27.2 ± 5.64% dissociation (Figure 20B). Lane 3 contained 12.5 nM of ALYREF-ΔC16, and resulted in 47.4 ± 4.52% dissociation. Lane 4 and 5 contained 50 nM of UAP56 and 6.25 nM and 12.5 nM ALYREF-ΔC16, respectively. Lane 4 had 86.7 ± 0.762% dissociation, and lane 5 had

87.8 ± 0.826% dissociation. These values are comparable to the percent dissociation observed by the same concentrations of ALYREF-WT. Lane 6 contained 6.25 nM of ALYREF-ΔRRM and resulted in a percent dissociation of 21.2 ± 2.67%. Lane 7 contained 12.5 nM of ALYREF-ΔRRM and had a percent dissociation of 36.2 ± 3.51%. These two concentrations were used in lanes 8 and 9 with 50 nM UAP56, which resulted in percent dissociations of 88.1 ± 0.332%, and 90.2 ± 0.535% respectively. These percentages are also comparable to the percentages recorded for ALYREF-WT at these concentrations with 50 nM UAP56.

The increase in R-loop dissociation for these two mutants could be explained in a couple of ways: the substrate used for R-loop flaps is unstable and too easily resolved, or ALYREF may undergo a conformation change that allows it to more accessibly bind to the RNA or UAP56. Regardless of the reason, the hypothesis that removing the C-terminal UAP56 binding motif would decrease its multiplicative effect on the R-loop dissociating activity of UAP56 was proven false. This indicates that another motif or other residues assist in binding to UAP56 as well as the documented binding UAP56 motif. This data also indicates that there may be other motifs that support RNA binding, and also perhaps that the mutant has a different conformation than the wildtype that allows it to more easily access RNA to bind. The crystallographic structure of ALYREF remains to be solved, with only the crystallographic structure of the RNA recognition motif having been documented.

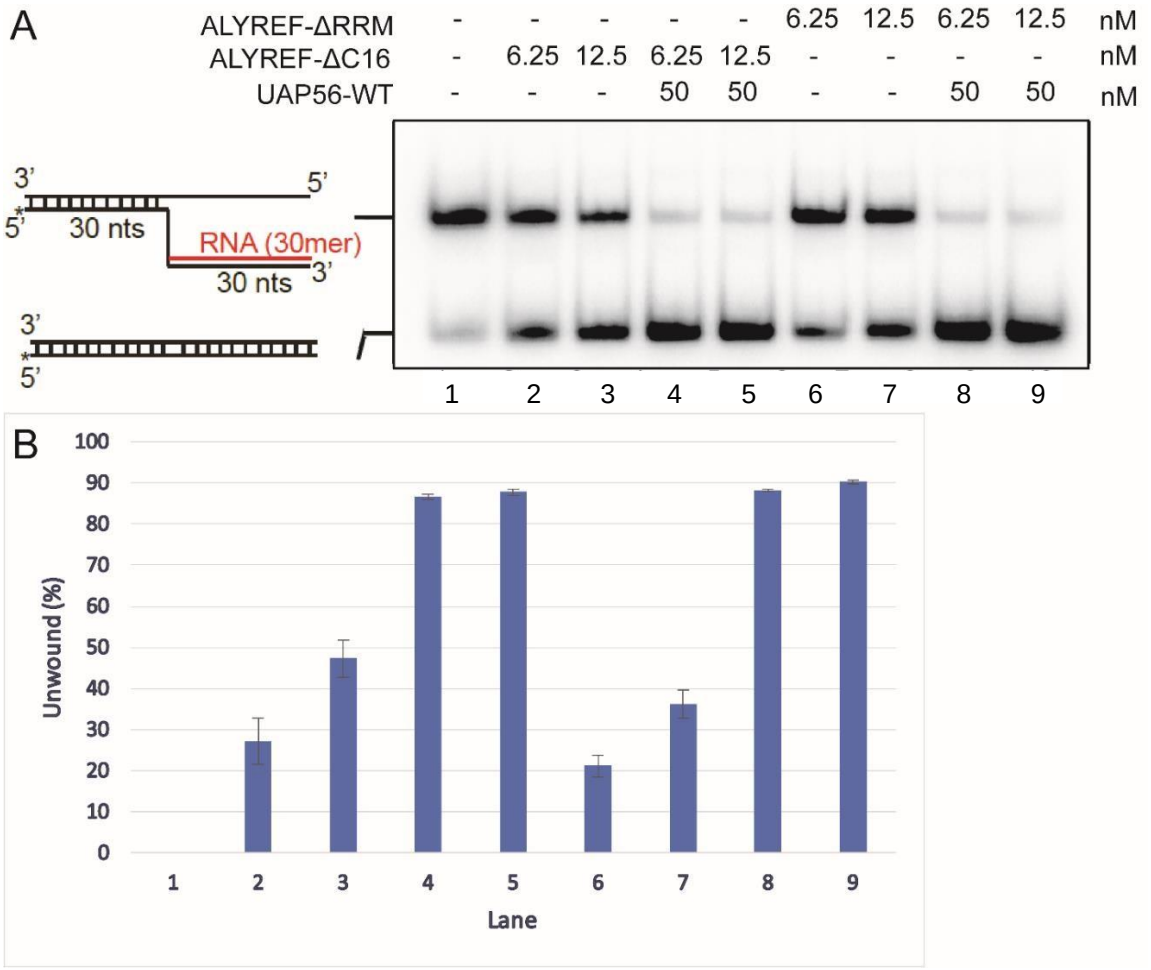

Figure 20. ALYREF-ΔC16 and ALYREF-ΔRRM titration with UAP56. A) The first lane shows the substrate alone. 50 nM UAP56 was used, as well as 6.25 nM and 12.5 nM ALYREF-ΔC16, and ALYREF-ΔRRM, as indicated. B) Graphical representation of the mean of the percentage of dsDNA product recovered after the reaction, with error bars showing standard deviation in percent (n=3).

EMSA with ALYREF-WT and ΔRRM

To discover the extent to which ALYREF-ΔRRM was able to bind to the substrate over the wildtype, two RNA binding assays were performed. These qualitative assays allow protein-RNA interactions to be observed. Longer streaks of smearing indicate larger complexes (RNA and protein) that are less mobile and not able to move easily through the gel. This smearing is only able to be detected because the 5' end of the R-loop substrate is labeled with ^{32}P-γ-ATP. Less smearing indicates the protein and

RNA do not interact, and therefore the labeled substrate is able to move through the gel more easily since it is not bound to a bulky protein.

For the first RNA binding assay, ALYREF-WT and ΔRRM were both diluted with K + 1000 mM KCl + 0.01% igepal + 1 mM DTT. Only 1 µL of each protein was used for each concentration, cumulatively 10 µL per reaction, which resulted in a final salt concentration 100 mM. The first lane was used as a control, which contained no ALYREF (Figure 21). The first three concentrations analyzed (12.5 nM, 25 nM, and 50 nM) did not appear to show any major differences in RNA binding. Due to the lack of differences, it was decided to proceed with higher salt concentration to reduce nonspecific binding due to charge interactions.

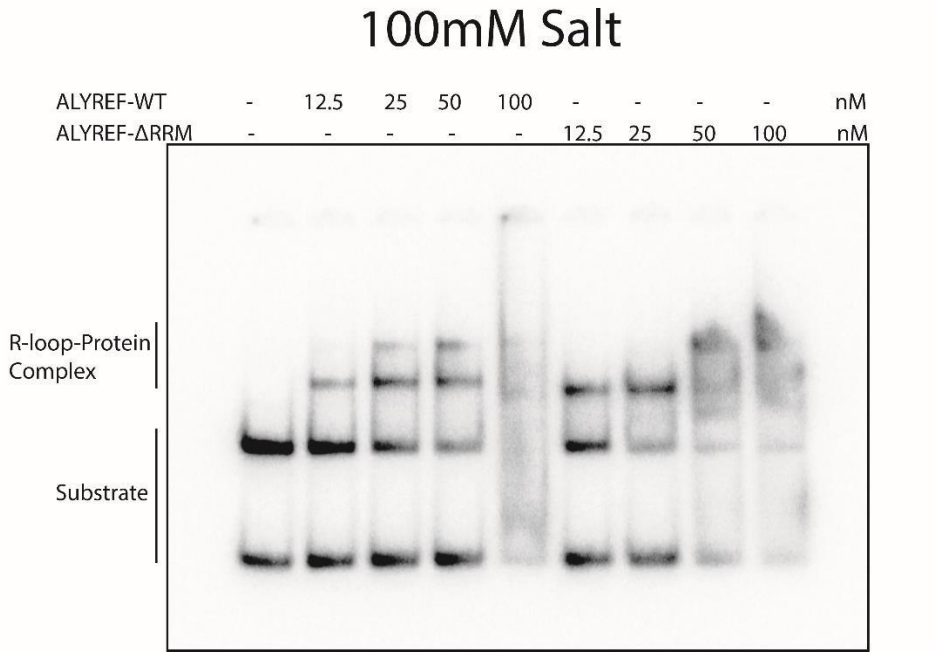

Figure 21. RNA binding assay with ALYREF-WT and ALYREF-ΔRRM at 100mM salt. RNA binding assays are useful for qualitatively observing RNA-protein interactions.

Salt concentration was increased by 50% for the next RNA binding assay, with a final concentration of 150 mM KCl. Increasing the salt concentration resulted in more obvious differences in binding (Figure 22). For the wildtype, only 100 nM appeared to bind at all to the substrate. In contrast, all concentrations of the mutant bound to the substrate, but there was no apparent increase in the intensity of the shifted band with the increase in protein concentration. This indicates that the increase in salt was successful in reducing non-specific binding, and that ALYREF-ΔRRM may bind to the substrate at an increased rate, as shown in Figure 20. However, a control assay should be conducted using substrate without RNA to confirm that ALYREF is not binding to DNA, and is specific for RNA.

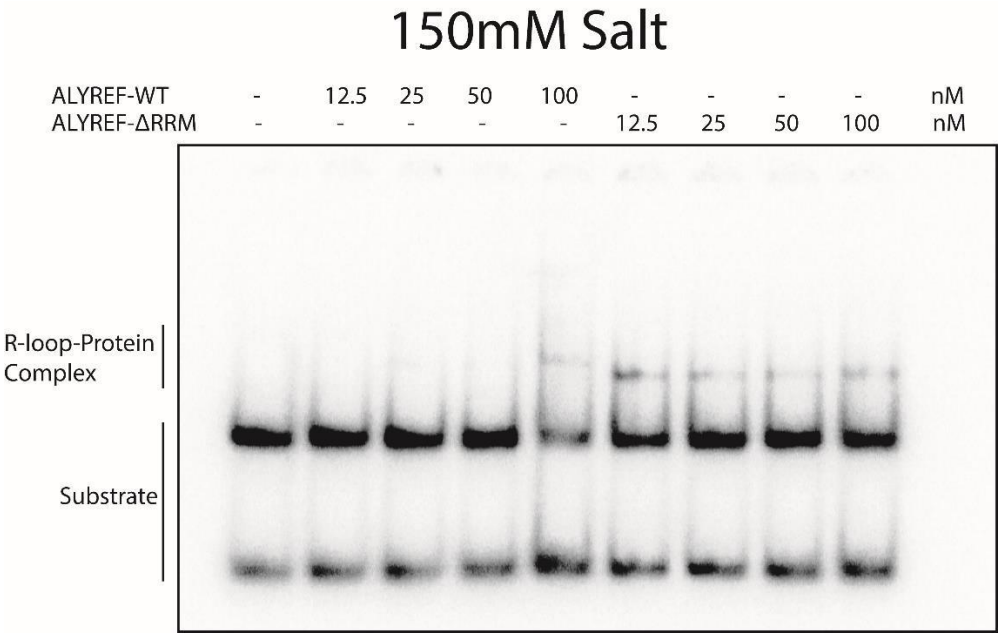

Figure 22. RNA binding assay with ALYREF-WT and ALYREF-ΔRRM at 150mM salt. Use of a higher salt concentration helped to prevent non-specific binding.

ALYREF-ΔNC expression

The data suggesting that the ALYREF-ΔC16 mutant still binds and also increases the ability of UAP56-WT to resolve R-loops prompted the development of another mutant to attempt to disrupt this. After searching through the literature, it was discovered that ALYREF also has an N-terminal UAP56 binding motif. With this new information, ALYREF-ΔNC was designed, which removed not only the C-terminal UAP56 binding motif, but also the N-terminal binding motif, presumably to create a mutant that would be unable to associate with UAP56. This mutant was purchased from Gene Universal in vector pET24a. Once received, the mutant was run on a 0.7% agarose gel, which revealed that the plasmid had a sufficient amount of supercoiled DNA, indicating that transformation into Rosetta cells for protein expression would be successful (Figure 23). The new mutant was compared to ALYREF-WT to confirm that it was the correct size.

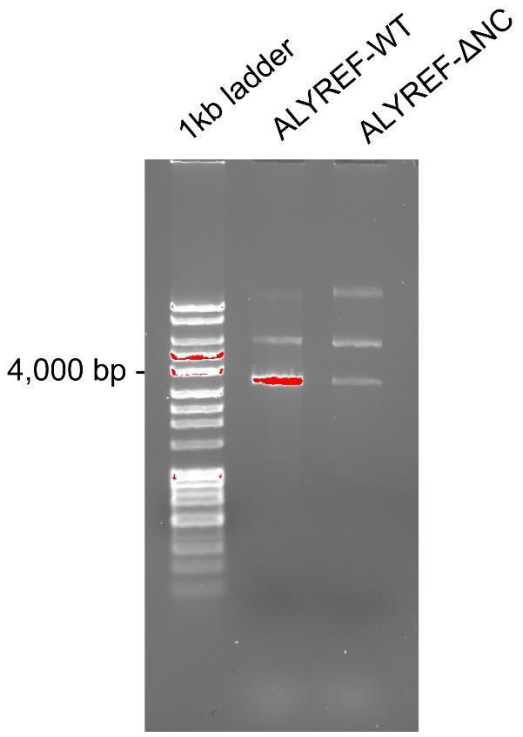

Figure 23. Agarose gel comparing ALYREF-WT and mutant ALYREF-ΔNC. The lower visible band shows supercoiled circular DNA.

After confirming the plasmid was correct, a transformation into Rosetta cells were performed. The transformation was plated with kanamycin and chloramphenicol, and incubated overnight at 37°C. The following day, a small-scale culture (120 mL) of each was started with antibiotics, and grown overnight at 37°C with shaking. The cell cultures were distributed into four flasks of 1 L portions of autoclaved LB with antibiotics for each mutant. The flasks were incubated at 37°C with shaking until OD_{600} reached 0.7-0.9. Then, protein expression was induced with 0.2 mM IPTG and incubated at 16°C overnight with shaking. The cells were then harvested the next day, and stored at -80°C.

ALYREF-ΔNC purification

The purification for ALYREF-ΔNC was started by breaking the cells in 100 mL of K + 150 mM KCl + 0.05% igepal + 1 mM DTT. Protease inhibitor tablets were added, as well as PMSF. The cell pellet was further broken down by stirring at 4°C, and sonication. The cell lysate was separated into supernatant and pellet by centrifugation, and sampled. The first column used to initially purify the protein was an SP Sepharose column. After the supernatant was loaded onto the column, the protein was eluted by HPLC, and sampled (Figure 24). The samples were analyzed by SDS-PAGE and Coomassie staining, and the gels for the first column revealed that although the mutant has a mass of about 23.7 kDa, the protein band appears to be located between 25 and 37 kDa. In total, 18 mL of elution fractions 20 through 24 were collected, combined and diluted to approximately 500 mM KCl.

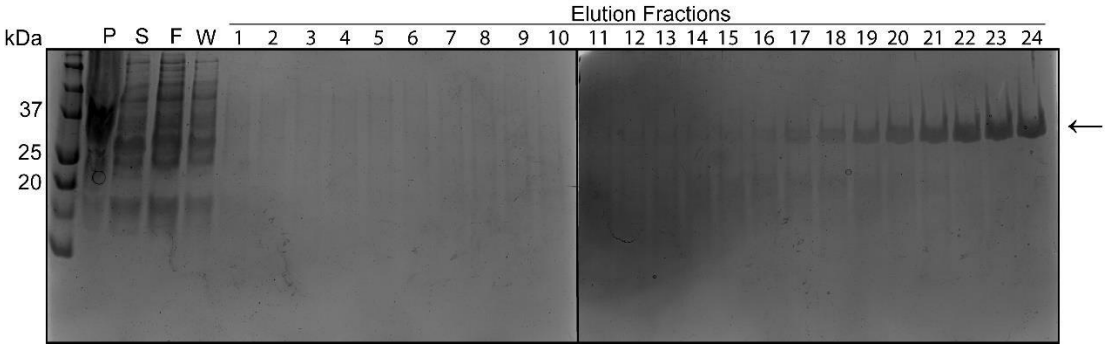

Figure 24. ALYREF-ΔNC affinity chromatography fractions. The cell lysate was purified using SP-Sepharose® Fast Flow affinity chromatography ("P", pellet, "S", supernatant, "F", flow-through, "W", wash). The fractions were separated by gel electrophoresis and stained with Coomassie Blue.

The second column used to further purify ALYREF-ΔNC was the HisTrap™. 21.6 mL of elution from the SP Sepharose column was loaded to the column by

Superloop. The binding buffer used was K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and the elution buffer was the binding buffer with 500 mM imidazole added. The protein was eluted with a linear gradient of 0-85% elution buffer. The elution fractions were sampled, and fractions 15 through 17 were collected since the SDS-PAGE analysis revealed they were the most pure and concentrated (Figure 25). The fractions were then combined, and mixed to a total of 5 mL. It was decided that the two columns were sufficient for the purification of this mutant, so the combined fractions needed to be prepared for storage by removing the high imidazole content. Concentration was performed using a 5 kDa MWCO centrifugal concentrator, changing the buffer to K + 500 mM KCl + 0.05% igepal + 1 mM DTT once after concentrating to 4 mL. Throughout this process, the protein solution seemed resistant to concentration, and would not concentrate below 5 mL after the addition of buffer. The buffer exchange still needed to be accomplished, so a decision was made to use a desalting column.

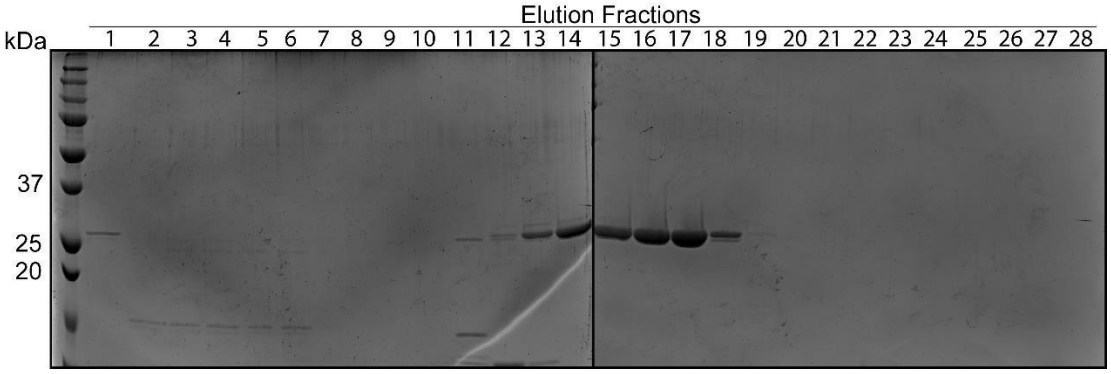

Figure 25. **ALYREF-ΔNC elution fractions from cation exchange chromatography.** The three most concentrated fractions were selected and pooled for further purification.

The 5 mL of elution from the cation exchange chromatography were separated into 2.5 mL portions, with one portion being set aside for later use. 2.5 mL of elution was combined with 2.5 mL of K + 500 mM KCl + 0.05% igepal + 1 mM DTT and

loaded to the desalting column in a 5 mL loop. The column was eluted with K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and the fractions were sampled and analyzed (Figure 26). Fractions 2 through 5 were collected, combined, and concentrated again with a 5 kDa MWCO centrifugal concentrator in 15 minute periods at 3900 RPM at 4°C. Concentrating proved difficult again, and although no aggregation was visible, concentration was stopped after reaching 3.0 mL to avoid any possible aggregation. 2.5 mL of the desalted protein elution were added to a 14 mL conical tube. The remaining 500 µL were aliquoted into 5 µL portions, and all were flash frozen with liquid nitrogen, and stored at -80°C.

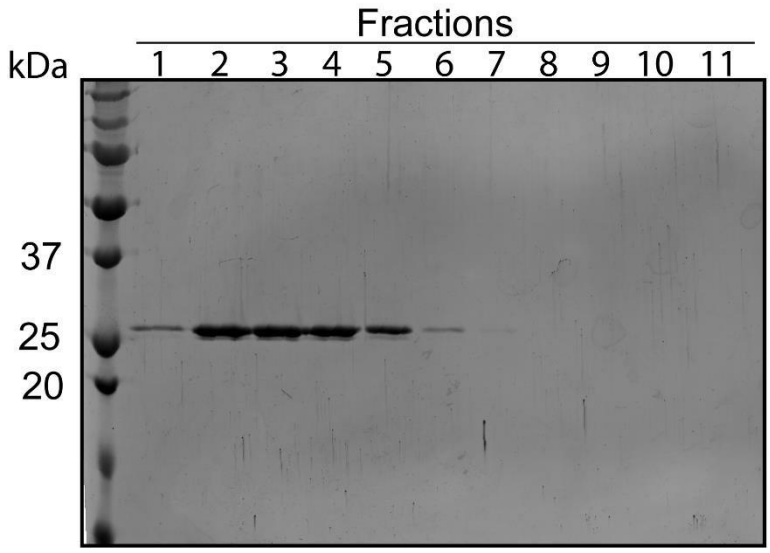

Figure 26. Desalting column fraction analysis by SDS-PAGE. ALYREF-ΔNC elutes near its molecular weight, using the 5 mL HiTrap™ Desalting column. The four most concentrated fractions were selected and pooled for further purification.

ALYREF-10E and ALYREF-17E expression

In an effort to create another mutant that is RNA binding deficient, ALYREF-10E and ALYREF-17E were designed. The first mutant has ten point mutations changing positively charged arginine residues to polar, negatively charged glutamic acid residues.

These mutations were selected specifically to disrupt RGG/RG motifs that are located heavily throughout ALYREF, and are known RNA binding segments. These ten point mutations were located throughout the protein. The second mutant, ALYREF-17E, contained the ten point mutations from the previous mutant and an additional seven point mutations that changed lysine, histidine, and more arginine residues into glutamic acid residues. Due to the large amount of charged amino acid replacements, the isoelectric points for these proteins were changed significantly. The wildtype protein has a theoretical pI of 11.05. ALYREF-10E and ALYREF-17E have theoretical pIs of 6.01 and 4.59, respectively.

Table 4. Representation of the mutants ALYREF-10E and ALYREF-17E and the amino acid substitutions used to create them.

Mutant	Point Mutations
ALYREF-10E	R27E, R30E, R34E, R36E, R38E, R224E, R227E, R231E, R233E, R235E
ALYREF-17E	R27E, R30E, R34E, R36E, R38E, K114E, K140E, K141E, R148E, H159E, R162E, K163E, R224E, R227E, R231E, R233E, R235E

These mutants were purchased from Gene Universal in pET24a vectors. 1 µL of each reconstituted mutant plasmid was sampled and run on a 0.7% agarose gel which revealed an acceptable amount of the DNA was in the supercoiled state (Figure 27). These new mutants were compared to ALYREF-WT to confirm that they are both the correct size. After confirming the plasmids were correct, transformations into Rosetta cells were performed. Each transformation was plated with kanamycin and chloramphenicol, and incubated overnight at 37°C. The following day, a small-scale culture (90 mL) of each was started with antibiotics, and grown overnight at 37°C with shaking. The cell cultures were distributed into three flasks of 1 L portions of autoclaved

LB with antibiotics for each mutant. The flasks were incubated at 37°C with shaking until OD_{600} reached 0.7-0.9. Then, protein expression was induced with 0.2 mM IPTG and incubated at 16°C overnight with shaking. The cells were then harvested the next day, and stored at -80°C.

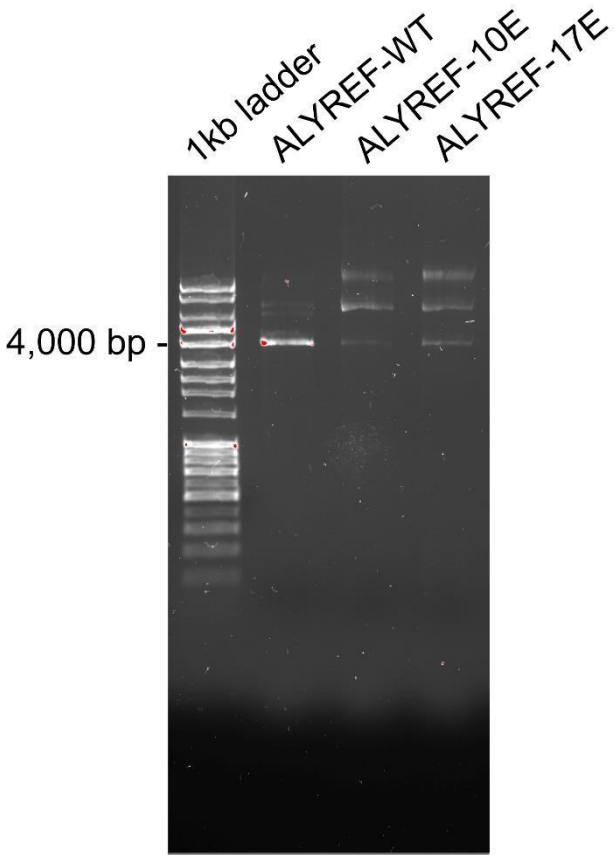

Figure 27. Agarose gel comparing ALYREF-WT and mutants ALYREF-10E and ALYREF-17E. The lowest visible band shows supercoiled circular DNA.

Purification of ALYREF-10E

The purification for ALYREF-10E was started by breaking the cells in 100 mL of K + 150 mM KCl + .05% igepal + 1 mM DTT. Protease inhibitor tablets were added, as well as PMSF. The cell pellet was further broken down by stirring at 4°C and sonication. The cell lysate was separated into supernatant and pellet by centrifugation, and sampled. The first column used to initially purify the protein was an SP Sepharose column. After

the supernatant was loaded onto the column, the protein was eluted by HPLC, and sampled. The flow through from the column still contained a large amount of protein. While neither gel appeared to contain a large amount of the mutant, some faint bands near the end of the second gel (lanes 17 through 21) were believed to be ALYREF-10E, and were collected (Figure 28). The flow through did appear to contain a larger amount of protein, which indicates it did not bind well to the SP Sepharose resin.

Figure 28. ALYREF-10E affinity chromatography fractions. The cell lysate was purified using SP-Sepharose® Fast Flow affinity chromatography ("L", ladder, "P", pellet, "S", supernatant, "F", flow-through, "W", wash). The fractions were separated by gel electrophoresis and stained with Coomassie Blue. ALYREF-10E elutes near 25 kDa, while its presumed weight is 27 kDa.

The second column used to further purify ALYREF-10E was the HisTrap™. 13.5 mL of elution from the SP Sepharose column was loaded to the column by Superloop. The binding buffer used was K + 500 mM KCl + 0.05% igepal + 1 mM DTT, and the elution buffer was the binding buffer with 500 mM imidazole added. The protein was eluted with a linear gradient of 0-85% elution buffer over 30 column volumes. The elution fractions were sampled, and the SDS-PAGE analysis revealed that ALYREF-10E did not appear to bind the column at all (Figure 29).

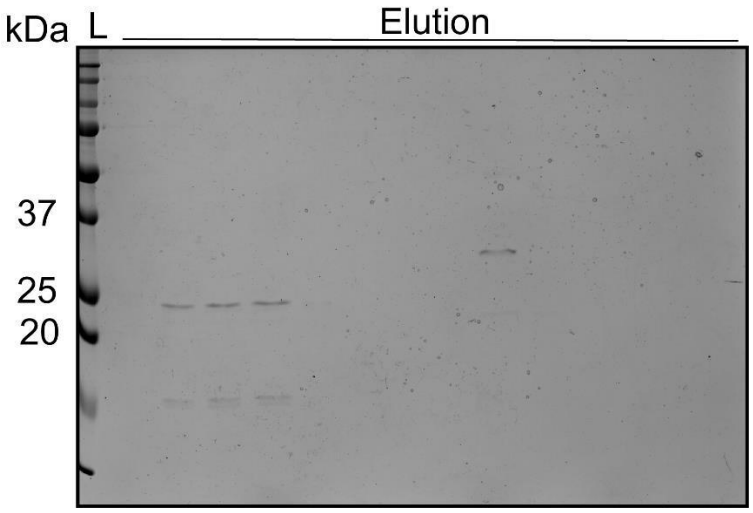

Figure 29. ALYREF-10E cation exchange chromatography fractions.
ALYREF-10E was expected to elute near 25 kDa but appeared to be present in a small amount, or not at all.

Since the results from the cation exchange column were not convincing, it was decided to perform a Western blot. The pellet, supernatant, SP column flow through, SP column wash, and previously collected elutions were sampled. The results of the blot showed the protein was expressed, but not eluted from the SP column whatsoever (Figure 30). After obtaining the Western blot results, it was decided to re-run the SP Sepharose flow through over a nickel column, which would then be eluted by HPLC.

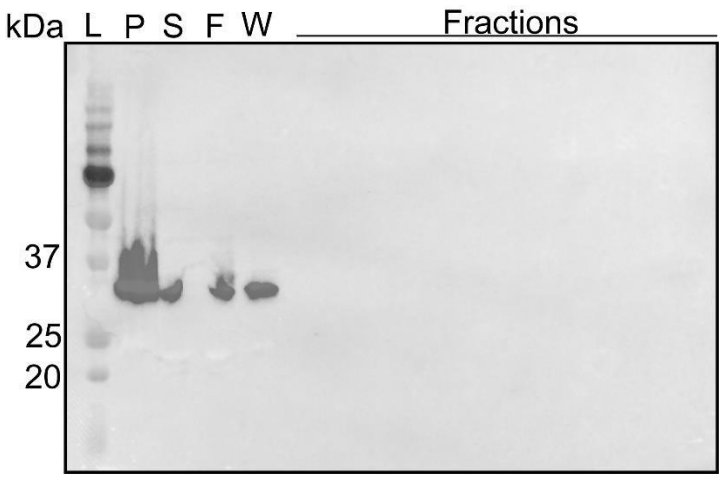

Figure 30. Western blot analysis of previous ALYREF-10E purification fractions. Multiple samples from the purification were selected and analyzed ("L", ladder, "P", pellet, "S", supernatant, "F", flow-through, "W", wash).

100 mL of the SP Sepharose flow through was added to the nickel column by the use of a pump. The binding buffer used was K + 150 mM KCl + 0.05% igepal + 1 mM DTT, and the elution buffer was K + 500 mM KCl + 0.05% igepal + 1 mM DTT + 500 mM imidazole. The elution fractions were sampled and analyzed by SDS-PAGE and Coomassie staining (Figure 31). The first three elution fractions sampled conveniently contained a large amount of the mutant without many other impurities. The fractions were then combined and mixed.

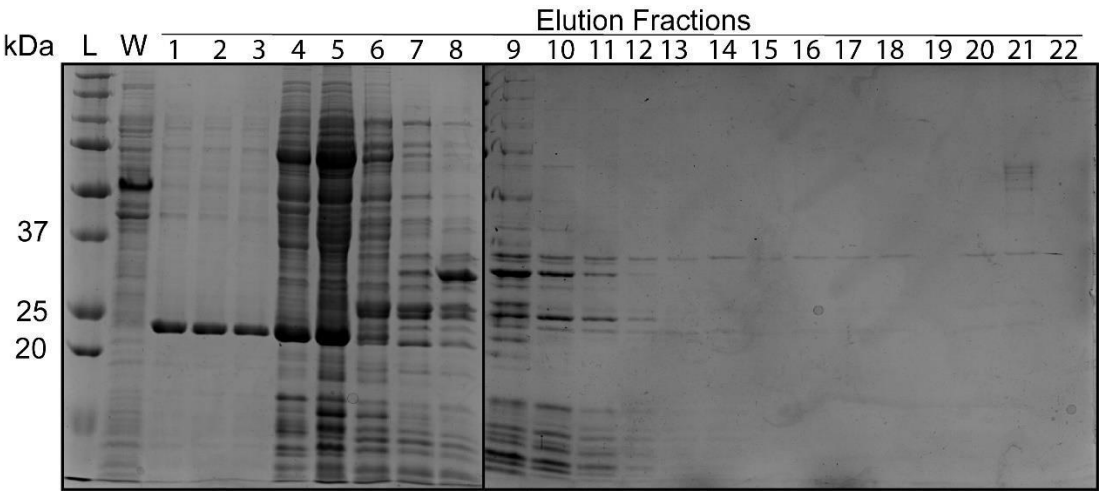

Figure 31. ALYREF-10E metal affinity chromatography fractions. The fractions were separated by gel electrophoresis and stained with Coomassie Blue. The first three elution fractions appear to have a lot of ALYREF protein, and were pooled to make a total amount of 12 mL.

Size exclusion chromatography was performed by loading the protein in 500 µL amounts for gel filtration using Superdex® 75 Increase and 500 mM K buffer. There was not a single significant UV peak using absorbance 280 nm, but the SDS-PAGE analysis of the fractions showed one fraction of purified ALYREF-10E (Figure 32). The single fraction was collected, and concentrated from 500 µL to 150 µL with a 5 kDa concentrator, aliquoted into 5 µL portions, and stored at -80°C.

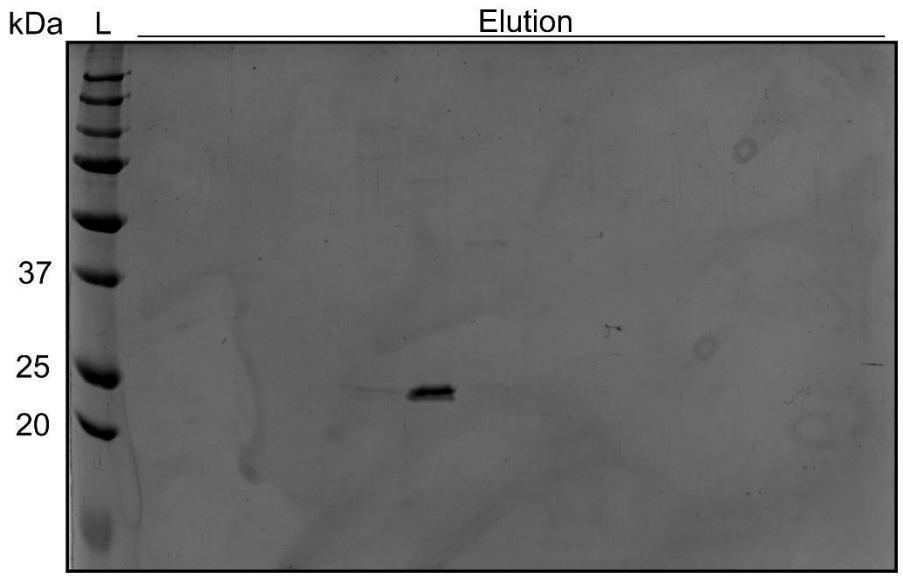

Figure 32. ALYREF-10E size exclusion chromatography fractions.
Size exclusion chromatography was performed twice with 500 µL portions for each run. The samples from each elution resulted in two identical SDS-PAGE, stained with Coomassie Blue. The only fraction containing protein was selected to be concentrated and aliquoted.

Unfortunately, ALYREF-17E was not able to be purified. After expressing in Rosetta cells, the purification was attempted twice with a variety of columns including nickel, Q Sepharose, His Trap, and Mono S. The first purification failed due to a series of mistakes and the inability of the protein to bind sufficiently to a column. The second purification attempt was deemed failed after the protein failed to bind the SP Sepharose column, and a Western blot was performed to assess if the correct protein was expressed. The Western blot detected no signal from the pellet, supernatant, flow through, or fractions, so the purification was put on hold in order to accomplish other experiments.

Concentration analysis of all purified proteins

An SDS-PAGE gel was run with bovine serum albumin standards of 0.5 µg and 1 µg, 0.5 µL and 1 µL of each ALYREF variant (Figure 33). The first three ALYREF variants' concentrations were previously determined. The concentrations for ALYREF-ΔNC and ALYREF-10E were determined to be roughly 250 µg/mL, and 150 µg/mL, respectively. This gel contains all ALYREF variants used for experiments.

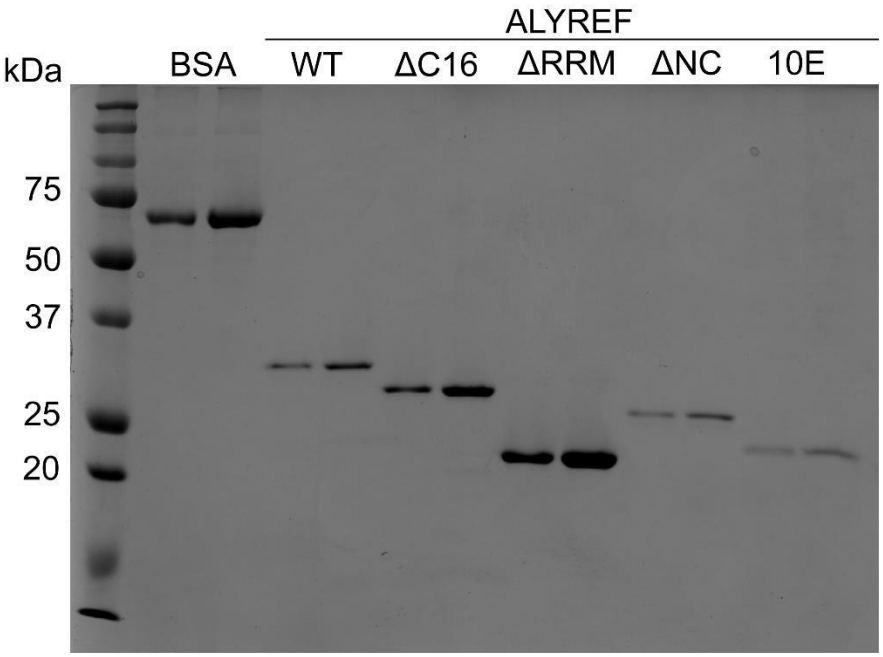

Figure 33. SDS-PAGE gel for all mutant concentration analysis.
ALYREF-WT, ALYREF-ΔC16, ALYREF-ΔRRM, ALYREF-ΔNC and ALYREF-10E were compared to bovine serum albumin standards.

ALYREF and UAP56 pull down assay

To test whether interaction between ALYREF-ΔNC and UAP56 was disrupted as hypothesized, pull-down assays were performed using nickel resin which ALYREF binds to. All assays were performed with a final concentration of 100 mM KCl. Purified ALYREF-WT and UAP56 were combined to incubate together and be sampled as a positive control. The second assay was performed without ALYREF, as a negative

control, assuming that UAP56 would elute entirely in the supernatant. The supernatant of each assay contained excess UAP56, and excess ALYREF that did not bind to the small amount of nickel resin used (Figure 34). The wash samples were taken after the fourth wash of the resin, and none showed any protein, meaning that the ALYREF did have a high affinity for the nickel resin and could not be knocked-off with low-salt low imidazole washes. The elutions consisted of 8 µL samples of the nickel resin combined with 20 µL of 2X SDS loading dye.

The first assay and second assays used as controls were consistent with previous findings, indicating that ALYREF-WT and UAP56 associate together, UAP56 does not interact with nickel beads, and that UAP56 is able to be pulled-down by ALYREF-WT (Figure 34). The assay with ALYREF-ΔNC showed a lot of excess UAP56 in the supernatant that did not bind with ALYREF. The wash contained no protein, but the elution contained ALYREF-ΔNC in a much larger amount than UAP56, indicating that mutant was not effectively able to bind to UAP56 as the wildtype protein can. Thus, the hypothesis that removing both the C-terminal and N-terminal UAP56 binding motifs in ALYREF could effectively disrupt its ability to bind to UAP56 was supported, but needs further testing. Another pull-down assay with higher salt should be used to prevent nonspecific binding at a higher rate, possibly reducing the small amount of UAP56 still bound to ALYREF.

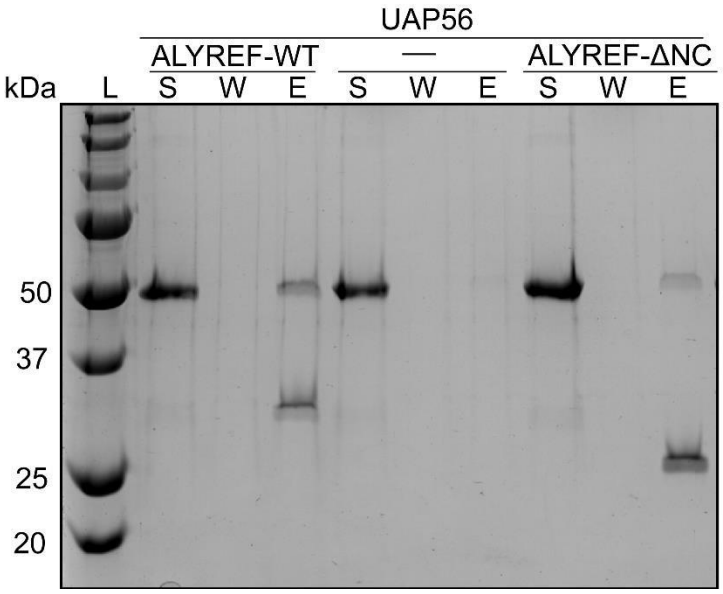

Figure 34. Pull down assay with ALYREF-WT, ALYREF-ΔNC, and UAP56-WT. The three lanes of each assay contain both ALYREF and UAP56, aside from the control which contains no ALYREF ("S", supernatant, "W", wash, "E" elution). Each assay is distinctly indicated by the ALYREF label separating them.

Testing of the R-loop resolution capabilities of UAP56 with ALYREF-ΔNC

After the discovery that ALYREF-ΔNC had a reduced ability to bind to UAP56, it was necessary to proceed with testing its effect on the R-loop resolution activity of UAP56. The lowest concentrations of ALYREF-WT that were originally tested, 6.25 nM and 12.5 nM, were used to test the mutant in comparison to the wildtype, and UAP56 continued to be tested with the standard concentration of 50 nM (Figure 35A). Lane 1 was used as a control with no added protein. It should be noted that the substrate used for this assay had higher than optimal background, so even lanes that appeared to have close to 100% dissociation were still quantified as having less than 80% (Figure 35B). Lanes 2 and 3 used ALYREF-WT concentrations of 6.25 nM and 12.5 nM, which resulted in dissociations of 4.97 ± 4.56%, and 10.0 ± 2.00%, respectively. Lane 3 used 6.25 nM

ALYREF-WT and 50 nm UAP56, and resulted in dissociation of 64.2 ± 3.42%. Lane 4 used 12.5 nM ALYREF-WT and 50 nM UAP56, and had dissociation of 72.5 ± 1.44%. Both of these percent dissociations are consistent with previous findings.

In testing ALYREF-ΔNC, both concentrations had similar percent dissociations with or without UAP56 present, which indicates any binding that may occur between the two proteins is not significant enough to affect the ability of UAP56 to resolve R-loops. Lane 6 with 6.25 nM ALYREF-ΔNC had a dissociation of 12.6 ± 5.59%, while lane 7 with 12.5 nM ALYREF-ΔNC dissociated at a rate of 33.8 ± 6.59%. This mutant also showed increased rates of dissociation in comparison to the wildtype, similar to mutants ALYREF-ΔC16 and ALYREF-ΔRRM. The addition of 50 nM UAP56 in lanes 8 and 9 resulted in dissociation of 10.7 ± 5.30%, and 40.1 ± 9.35%, respectively. These values show significantly less dissociation for ALYREF-ΔNC and UAP56 than the dissociation the wildtype and UAP56 have. This data emphasizes that ALYREF has a direct effect on the DNA/RNA helicase and R-loop dissociation activities of UAP56.

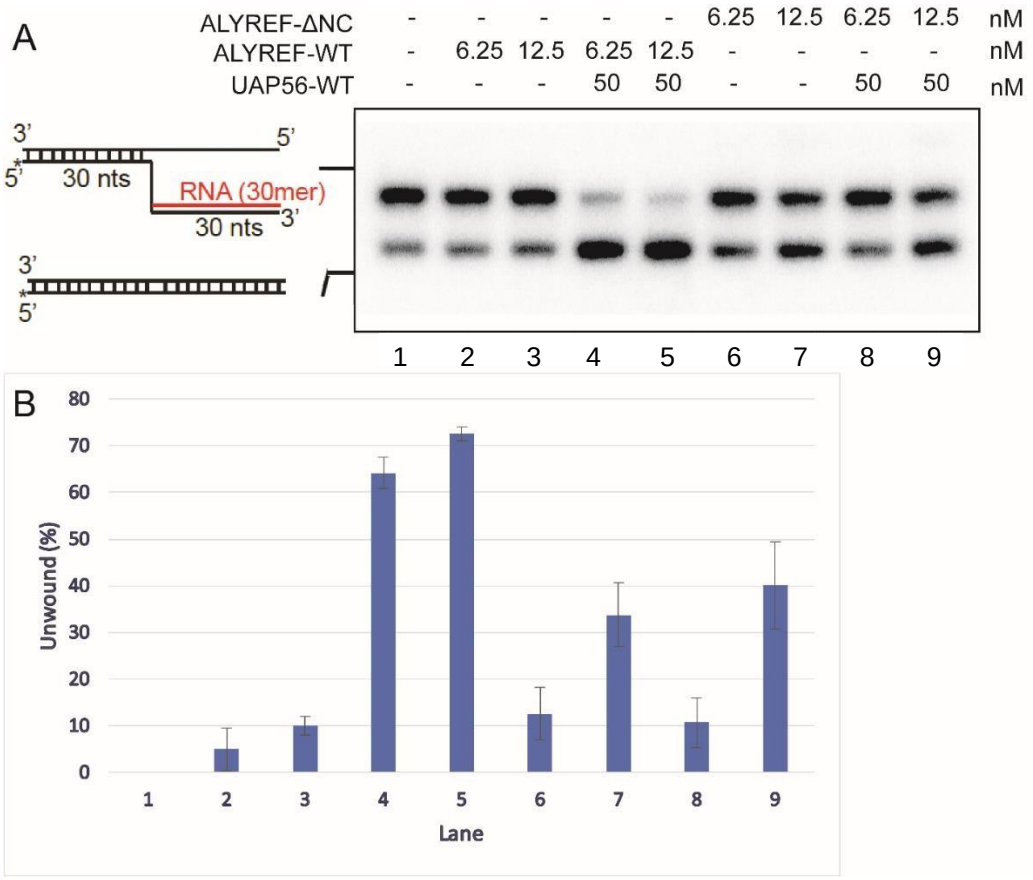

Figure 35. ALYREF-WT and ALYREF-ΔNC titration comparison with UAP56-WT. A) The first lane shows the background of the substrate. 50 nM UAP56 was used, as well as 6.25 nM and 12.5 nM ALYREF-ΔNC and ALYREF-WT. B) Graphical representation of unwound percentage for each lane from the assays, with error bars showing standard deviation in percent (n=3).

IV. SUMMARY AND CONCLUSIONS

The major goals of this project were to characterize the ability of UAP56 and its mutants to resolve RNA:DNA hybrids, and to identify the effect of ALYREF on the R-loop capabilities of UAP56 *in vitro* and in cells. We purified human UAP56-WT, UAP56-E197A, UAP56-K95A, ALYREF-WT, ALYREF-ΔC16, ALYREF-ΔRRM, ALYREF-ΔNC, and ALYREF-10E to accomplish this, and test the effect further. It was established through R-loop resolution assays that UAP56-WT does resolve R-loop flaps at a rate of 80.8 ± 1.55% at a high 200 nM concentration. However, the wildtype protein's activity appeared to decrease significantly at lower concentrations (Figure 8). The UAP56-K95A mutant had negligible resolution activity, which supports the hypothesis that the Lys-95 residue is necessary for helicase activity because it binds ATP [31]. Initial testing of UAP56-E197A showed it may have RNase contamination and it needs to be expressed and purified again.

The Effect of ALYREF on UAP56

ALYREF and four of its mutants were expressed and purified to identify and explore the effect of ALYREF on the R-loop resolution capabilities of UAP56. The interactions between ALYREF-WT and UAP56-WT and its mutants were confirmed by pull down assay before testing in R-loop resolution assays. Data from the R-loop resolution assays indicated that ALYREF has a stimulatory effect on the ability of UAP56 to resolve R-loops. 12.5 nM of ALYREF alone had dissociation that was observed at 10.3 ± 4.15%. 50 nM UAP56 and resulted in 16.1 ± 2.88% dissociation, while a reaction mixture containing 50 nM UAP56 and 12.5 nM ALYREF had dissociation that was observed at 88.3 ± 2.90% (Figure 10). Through similar assays,

ALYREF-WT continued to appear to multiplicatively increase the DNA/RNA helicase and R-loop dissociation activities of UAP56.

One of the ALYREF mutants tested, ALYREF-ΔC16, was designed to test the hypothesis that disrupting the interaction between ALYREF and UAP56 would result in decrease of R-loop resolution by UAP56. The interaction between the two proteins was thought to be sufficiently interrupted by deletion of the C-terminal UAP56 binding motif on ALYREF, as indicated by co-immunoprecipitation data [32]. The interaction between ALYREF-ΔC16 and UAP56 was tested by pull down assay, which revealed that interaction was maintained (Figure 19). The difference between our results and the previously published data may be due to the interaction of other proteins in the HeLa cell extracts used immunoprecipitation, whereas our proteins were isolated and purified to near homogeneity. The cell extracts used in immunoprecipitation may have contained other proteins that further inhibited the interaction between ALYREF and UAP56, while these other protein interactions did not exist in our assays. After the discovery that the interaction was not disrupted, the effect of the mutant on UAP56 was analyzed by R-loop resolution assays. 50 nM UAP56 and 12.5 nM ALYREF-ΔC16 had 87.8 ± 0.826% dissociation, which is comparable to the percent dissociation observed by the same concentrations of ALYREF-WT (Figure 20). This data indicated that another UAP56binding region must be present, and it is sufficient to mediate the interaction with UAP56 and to stimulate the R-loop resolution activity by UAP56.

Further exploration of the literature confirmed there are two short, conserved motifs that mediate interaction with UAP56 [32]. LDxxLD is the motif that was affected in the C-terminal UAP56 binding motif deletion. The second motif, SLDD/E is located

in the N-terminal region of ALYREF. So, another mutant, ALYREF-ΔNC, with deletions of both the N and C termini was designed. After the expression and purification of this mutant, the interaction with UAP56 was tested in a pull down assay (Figure 34). The assay revealed a substantially lower amount of UAP56 in the elution, which indicates binding was impaired to an extent, although there was some weak interaction between UAP56 and ALYREF-ΔNC. In the future, we can determine if the weak interaction between UAP56 and ALYREF-ΔNC would be disrupted by increasing the stringency of the pulldown condition (e.g. higher salt concentration). Assessing the effects of this mutant by R-loop resolution assay revealed that 12.5 nM ALYREF-ΔNC had 33.8 ± 6.59% dissociation, and the addition of 50 nM UAP56 resulted in 40.1 ± 9.35% dissociation (Figure 35). These values are not statistically different, and certainly have lower percent dissociation in comparison to UAP56 and ALYREF-WT together, but more work needs to be done to definitively answer if this mutant is capable of preventing ALYREF from associating with UAP56.

In order to test whether the ability of ALYREF to bind to RNA has an effect on its ability to increase the DNA/RNA helicase and R-loop dissociation activities of UAP56, an RNA binding defect truncation mutant, ALYREF-ΔRRM, was designed. After the expression and purification of this mutant, the interaction with UAP56 was confirmed with a pull down assay, in which proportional amounts of ALYREF and UAP56 were present in the elution (Figure 19). This mutant was tested with R-loop resolution assays, which revealed that 12.5 nM ALYREF-ΔRRM with 50 nM UAP56 resulted in 90.2 ± 0.535% dissociation (Figure 20). This percent dissociation is similar to that recorded for 12.5 nM ALYREF-WT with 50 nM UAP56. The hypothesis that

ALYREF-ΔRRM would exhibit decreased R-loop dissociation ability with UAP56 in relation to ALYREF-WT needs further testing to conclude. An interesting finding in this assay was that the RRM truncated mutant appeared to dissociate R-loops at least twice the rate that ALYREF-WT did in previous R-loop resolution assays.

After the unexpected results from the R-loop resolution assays, the extent to which ALYREF-ΔRRM binds to the substrate prompted us to run an EMSA to qualitatively observe the RNA-protein interactions. The first assay at 100 mM KCl concentration showed few differences between the wildtype and mutant at concentrations of 12.5, 25, 50, and 100 nM (Figure 21). A second EMSA was performed at higher salt (150 mM) to reduce non-specific binding. In this assay, 100 nM of ALYREF-WT was the only concentration that appeared to bind at all to the substrate, while all concentrations (12.5, 25, 50, 100 nM) of the mutant bound to the substrate, but with no increase in shifted band intensity as concentration of the protein increased (Figure 22). This is consistent with the results from the R-loop resolution assay that indicated ALYREF-ΔRRM does actually bind to the substrate at an increased rate, which could be due to a conformation change which allows better contact with RNA. These results also indicated that there are other motifs outside of the RRM that contribute to RNA binding.

Further research into RNA binding motifs revealed that multiple RGG/RG motifs can form high-affinity interactions with RNA structures [33]. These motifs are typically located in disordered regions, of which ALYREF has two. On either side of the RRM, there is a flanking disordered region, and these RGG/RG motifs are located heavily throughout these regions of ALYREF. For the design of the next ALYREF mutant, the arginine of each of these motifs were selected for mutation to glutamic acid, reducing the

interaction between negatively charged molecules (RNA and DNA). ALYREF-10E was the first mutant designed by targeting these RRG/RG motifs, with a total of 10 point mutations of arginine to glutamic acid (Table 4). A second mutant, ALYREF-17E was designed as well, with the point mutations of the 10E mutant and an additional seven point mutations of positively charged lysine and histidine residues to glutamic acid as well. ALREF-10E was able to be expressed and purified, but there were many issues with the purification of ALYREF-17E because it did not bind well to any of the columns used.

Future Work

The pull down assay testing the interaction between ALYREF-ΔNC and UAP56 needs to be repeated using higher salt concentration to reduce non-specific binding. The interaction between ALYREF-10E and UAP56 still needs to be confirmed, and RNA binding assays should be performed to assess the ability of the mutant to bind to negatively charged molecules. ALYREF-17E will need to be expressed and purified, and tested as ALYREF-10E is to assess if the binding to negatively charged molecules is affected with either mutant, or to what extent it is affected. If either of these mutants show decreased binding to the substrate, they should be tested in R-loop resolution assays alongside UAP56.

Our collaborator, Andres Aguilera, will assist in testing the effects of our mutants on R-loop resolution in cells. It would also be interesting to find out the concentrations at which these proteins are normally present in cells, so realistic concentrations can also be used to test the effects *in vitro*. While we have established the ability of ALYREF to upregulate the DNA/RNA helicase and R-loop dissociation activities of UAP56, further

investigation on the extent and mechanism of this regulation by ALYREF needs to be done to flesh out the implications this may have on potential cancer therapies relating to the TREX complex.

REFERENCES

1. Prevention, C.f.D.C.a., NCHS - Leading Causes of Death: United States | Data. Centers for Disease Control and Prevention.

2. Centers for Disease Control and Prevention (CDC), in Encyclopedia of Public Health. Springer Netherlands. p. 105-105.

3. Richard, P. and J.L. Manley, R Loops and Links to Human Disease. Journal of Molecular Biology, 2017. **429**(21): p. 3168-3180.

4. J., et al., Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. Molecular Cell, 2014. **56**(6): p. 777-785.

5. Aguilera, A. and T. García-Muse, R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular Cell, 2012. **46**(2): p. 115-124.

6. Skourti-Stathaki, K., et al. , R-Loops Induce Repressive Chromatin Marks over Mammalian Gene Terminators. Nature, 2014. **vol. 516**: p. pp. 436-439.

7. Yu, K., et al., R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol, 2003. **4**(5): p. 442-51.

8. James, P.A. and K. Talbot, The molecular genetics of non-ALS motor neuron diseases. Biochim Biophys Acta, 2006. **1762**(11-12): p. 986-1000.

9. Palau, F. and C. Espinos, Autosomal recessive cerebellar ataxias. Orphanet J Rare Dis, 2006. **1**: p. 47.

10. Mehta, A. and J.E. Haber, Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. Cold Spring Harbor Perspectives in Biology, 2014. **6**(9): p. a016428-a016428.

11. Li, X. and J.L. Manley, Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. Cell, 2005. **122**(3): p. 365-378.

12. Skourti-Stathaki, K., N.J. Proudfoot, and N. Gromak, Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2dependent termination. Mol Cell, 2011. **42**(6): p. 794-805.

13. Sollier, J., et al., Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell, 2014. **56**(6): p. 777-85.

14. Santos-Pereira, J.M. and A. Aguilera, R loops: new modulators of genome dynamics and function. Nature Reviews Genetics, 2015. **16**(10): p. 583-597.

15. Huertas, P. and A. Aguilera, Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. Molecular Cell, 2003. **12**(3): p. 711-721.

16. Skourti-Stathaki, K. and N.J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & Development, 2014. **28**(13): p. 1384-1396.

17. Cristini, A., et al., RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Reports, 2018. **23**(6): p. 1891-1905.

18. Hodroj, D., K. Serhal, and D. Maiorano, Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. Nucleus, 2017. **8**(5): p. 489-495.

19. Sridhara, S.C., et al., Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. Cell Reports, 2017. **18**(2): p. 334-343.

20. Bhatia, V., et al., BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature, 2014. **511**(7509): p. 362-365.

21. García-Rubio, M.L., et al., The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLOS Genetics, 2015. **11**(11): p. e1005674.

22. Hatchi, E., et al., BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Molecular Cell, 2015. **57**(4): p. 636647.

23. Chang, E., and Peter Stirling, Replication Fork Protection Factors Controlling RLoop Bypass and Suppression. Genes & Development, 2017. **8**(1): p. 33.

24. Sträßer, K., et al., TREX is a conserved complex coupling transcription with messenger RNA export. Nature, 2002. **417**(6886): p. 304-308.

25. Katahira, J., mRNA export and the TREX complex. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2012. **1819**(6): p. 507-513.

26. Domínguez-Sánchez, M.S., et al., Genome Instability and Transcription Elongation Impairment in Human Cells Depleted of THO/TREX. PLoS Genetics, 2011. **7**(12): p. e1002386.

27. Heath, C.G., N. Viphakone, and S.A. Wilson, The role of TREX in gene expression and disease. Biochemical Journal, 2016. **473**(19): p. 2911-2935.

28. Domínguez-Sánchez, M.S., et al., Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers. BMC Cancer, 2011. **11**(1).

29. Saito, Y., et al., ALY as a potential contributor to metastasis in human oral squamous cell carcinoma. Journal of Cancer Research and Clinical Oncology, 2012. **139**(4): p. 585-594.

30. Shen, C., et al., TOR Signaling Is a Determinant of Cell Survival in Response to DNA Damage. Molecular and Cellular Biology, 2007. **27**(20): p. 7007-7017.

31. Shen, J., L. Zhang, and R. Zhao, Biochemical Characterization of the ATPase and Helicase Activity of UAP56, an Essential Pre-mRNA Splicing and mRNA Export Factor. Journal of Biological Chemistry, 2007. **282**: p. 22544-22550.

32. Gromadzka, Agnieszka M., et al., A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. Nucleic acids research. 2016. **44**(5): p. 2348-2361.

33. Chong, P. A., et al., RGG/RG Motif Regions in RNA Binding and Phase Separation. Journal of Molecular Biology, 2018. **430**(23): p. 4650-4665.